Exploration of chemoprotective activities in *Arabidopsis thaliana* activation tagged lines

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Veronika Ungewickell

aus Göttingen

Köln 2012

Berichterstatter: Prof. Dr. U.-I. Flügge

Prof. Dr. S. Waffenschmidt

Tag der mündlichen Prüfung: 14.06.2012

What is a scientist after all? It is a curious man looking through a keyhole, the keyhole of nature, trying to know what's going on.

Jacques Yves Cousteau (1910-1997)

Table of Contents

1 Introduction

The essential metabolic process that occurs in all plant cells is defined as the primary metabolism (*1*). In contrast, secondary metabolites are organic compounds that are not necessary for growth, development, or reproduction but are important for the plants survival under advert environmental conditions (*1,2*). Plants produce secondary metabolites in response to abiotic and biotic environmental stress. So far more than 100,000 secondary compounds have been identified in plants (*3*). In *Arabidopsis thaliana (A. thaliana)* these fall into seven major classes: Indole and indole-sulfur compounds, glucosinolates, phenylpropanoids, benzenoids, flavonoids, terpenes, and fatty acid derivatives (*4*) (see Fig. 1 for a schematic summary of the biosynthetic pathways). Many of the compounds are derived from related pathways. Some metabolites - termed phytochemicals - have been shown to have beneficial effects for humans, which include anti-cancerogenic and anti–mutagenic compounds (*5*). Due to these effects, Sporn (1976) and Wattenberg (1985) introduced the idea of chemoprevention. The aim is to intervene with cancer development by prevention, which includes the uptake of dietary phytochemicals (*6,7*). It is of great interest today to identify novel natural occurring compounds with chemoprotective properties and to elucidate their functional mechanism as preventive agents against cancer (*8-10*). Many compounds have been identified in plants with the desired properties. Nevertheless, the identified substances are only a fraction of the predicted number of compounds in plants, which is estimated between 200,000 to 1,000,000 metabolites (*11*). In the following chapters, the function of secondary metabolites in plants and their chemopreventive properties for mammals will be summarized.

One way of function of phytochemicals is via the induction of phase II and detoxification enzymes that are mediated by electrophile response elements (EpRE) located in the promotor region of the respective genes. This mechanism will be described and the use of EpRE as a reporter system for the detection of chemopreventive agents will be summarized in more detail. Current strategies for untargeted metabolite profiling will also be presented. The TAMARA (**t**ransposable element mediated **a**ctivation tagging **m**utagenesis in **Ara**bidopsis) population of *A. thaliana* plants, which are analyzed in this work, will be introduced.

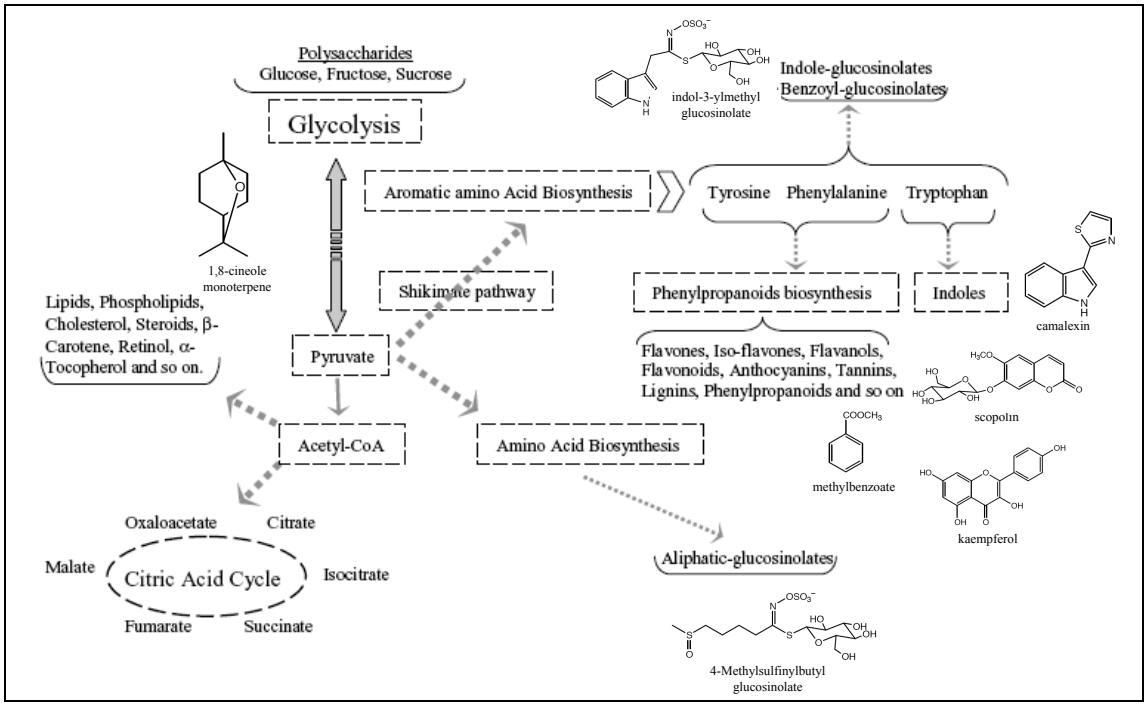


Fig. 1: General biosynthetic pathway of Brassicacae metabolites. Structures of representative compounds are given. Adapted from (*4,12*).

1.1 Secondary metabolites in Arabidopsis thaliana

1.1.1 Indole and indole-sulfur compounds

Indole and indole-sulfur compounds are derived from the amino acid tryptophan. One prominent representative of this group is the phytoalexin camalexin (3-thiazol-2'-ylindole). Other tryptophan-derived compounds are indole glucosinolates and simple tryptophan derivatives (*4*). The biosynthesis pathways of indole glucosinolates, camalexin and auxin (indole-3-acetic acid (IAA)) are closely related as shown in Fig. 2 (*4,13*). The enzymes CYP79B2 and CYP79B3 convert tryptophan into indole-3-acetaldoxime, which is dehydrated to indole-3-acetonitrile (IAN). Dihydrocamalexic acid is formed in the next step. This compound is decarboxylated by PAD3 (CYP71B15, phytoalexin deficient 3), resulting in the production of camalexin. The phytohormone auxin is either formed from tryptophan or upon the accumulation of IAN. IAN is then hydrolyzed by nitrilases to form auxin. Indole glucosinolates are, like camalexin, derived from indole-3-acetaldoxime (*13*). The general role of glucosinolates is described in chapter 1.1.2. Camalexin is involved in plant defense responses against various biotrophic and necrotrophic pathogens such as bacteria,

viruses, fungi and oomycetes (*4,14-17*). Upon pathogen attack and by abiotic reactive oxygen species (ROS) generating elicitors such as silver nitrate, camalexin biosynthesis is induced (*2,16,18,19*). The beneficial effect of camalexin for human health was recently shown by Mezencev *et al*. (2011), who could demonstrate that camalexin induces apoptosis specifically in Jurkat (T-cell leukemia) cells (*20*).

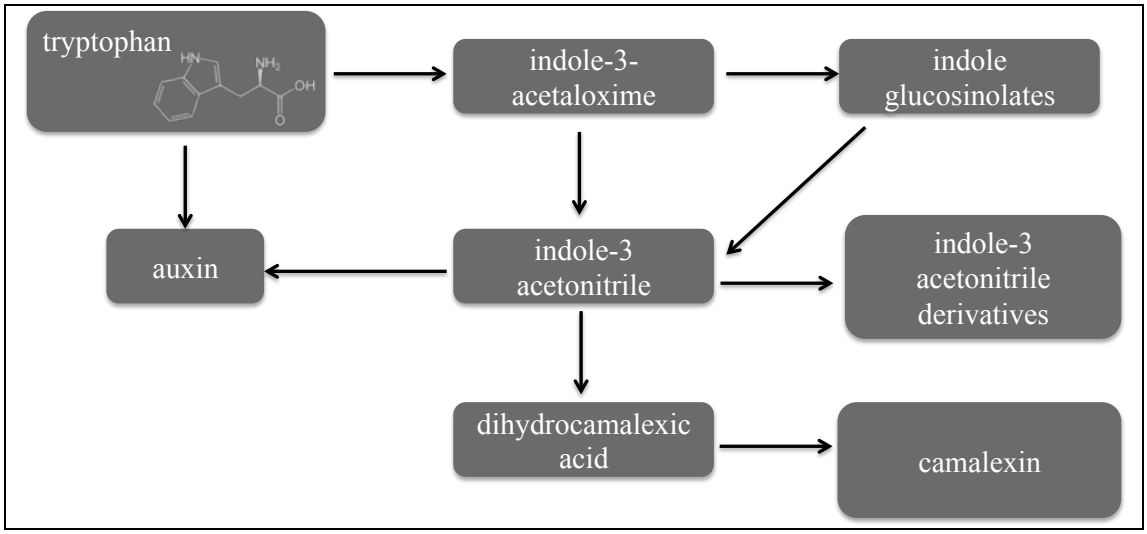


Fig. 2: Schematic diagram of camalexin biosynthesis from tryptophan. The pathway is closely related to the synthesis of auxin and indole glucosinolates due to the common precursors tryptophan, indole-3-acetaloxime and indole-3 acetonitrile (IAN). Adapted from (*13*).

1.1.2 Glucosinolates

Glucosinolates are amino acid derived nitrogen- and sulfur-containing compounds that are mainly found in species of the *Capparales* order. Around 200 different structures are presently known from this group of compounds, 34 glucosinolate forms have been detected in seeds and leaves of various *A. thaliana* ecotypes (*21*). Glucosinolates are activated by myrosinases, which are enzymes that cleave the thioglycosidic bond, releasing an unstable aglycone. The aglycone spontaneously rearranges and forms the toxic and active isothiocyanate (ITC) (*22*). The myrosinases are stored in myrosin cells (idioblasts), separately from glucosinolates (*23,24*). Upon tissue damage, caused for example by chewing insects or processing of foods, the enzymes and glucosinolates come in contact with each other. This leads to the generation of the active compounds. Depending on the type of aglycone, the reaction temperature, the pH and the presence of ferrous ions, the products of hydrolysis can vary. Nitriles, thiocyanates, epithionitriles and oxazolidine-2-thiones are formed (*22,25*). This is also controlled by the presence of protein factors such as nitrile-specifier proteins or epithio specifier proteins (*25-27*) The glucosinolate-myrosinase

system is often referred to as the ‘mustard oil bomb’ (Fig. 3) (*28*). Glucosinolate hydrolysis products function as repellents against a variety of organisms such as insects, nematodes, bacteria and fungi, though they can also be an attractant to some insects (*29-38*). The amino acids alanine, valine, leucine, isoleucine, phenylalanine, methionine, tyrosine and tryptophan are the basis for glucosinolate biosynthesis. They can be grouped depending on the amino acid, into aliphatic, indole and aromatic glucosinolates (Fig. 2 shows that indole glucosinolates are derived from tryptophan). The biosynthesis of glucosinolates is divided into the three stages: Side chain elongation, core structure development and secondary side chain modifications. During the first step, one or more methylene groups extend certain aromatic and aliphatic amino acids. Then the core structure of the glucosinolate is formed via rearrangement of the molecule. The side chains of the glucosinolate molecule are then further modified. During the biosynthesis of glucosinolates, several precursor compounds are formed: *N*-hydroxy amino acids, aldoximes, thiohydroxymic acids and desulfoglucosinolates. Cytochrome-P450-dependent mono-oxygenases (CYPs) catalyze the production of aldoximes from amino acids. The elongation step and the modifications of the side chains are the basis of the structural diversity of glucosinolates (*39,40*). The consumption of cruciferous vegetables, which have a high content of glucosinolates, reduces the risk of cancer mainly due to its breakdown products (*26,41-45*). This could be demonstrated in studies on bladder and prostate cancers but also in lung cancer and cancer in the gastrointestinal tract (*42,46-48*). Especially the ITC sulforaphane was shown to be an effective component against tumorigenesis (*44,49-52*). Evidence exists that also intact glucosinolates have chemopreventive activity, as shown with glucoraphanin (4-Methylsulfinylbutyl glucosinolate) and glucoerucin (4-Methylthiobutyl glucosinolate) (*53*). In comparison to the available data about ITCs, little is known about the chemoprotective effect of other glucosinolate breakdown products such as thiocyanates, nitriles and epithionitriles (*26*). Though, it could recently be demonstrated that the epithionitriles 1-cyano-2,3-epithiopropane, 1-cyano-3,4-epithiobutane and 1-cyano-4,5-epithiopentane induce phase II enzymes in rat liver epithelial cells (*54*). Also the nitrile crambene (1-cyano-2-hydroxy-3-butene), which is a breakdown product of the aliphatic glucosinolate progoitrin, has been demonstrated to induce phase II enzymes, especially when administered in combination with indole-3-carbinol (I3C) (*55-59*).

The most common indole glucosinolates are indol-3-ylmethylglucosinolate (glucobrassicin, I3M), 1-methoxy-I3M (1MO-I3M), 4-methoxy-I3M (4MO-I3M) and 4-hydroxy-I3M (4HO-I3M). The main indole glucosinolate in leaves is glucobrassicin. In roots, the concentration of indole glucosinolates is particularily high. The hydrolysis of indole glucosinolates also results in the formation of different breakdown products. Unlike ITCs derived from aliphatic or aromatic glucosinolates, the indole glucosinolate derived ITCs are unstable. To date the only indole derived ITC detected is 1-methoxyindol-3-ylmethyl-ITC, which is derived from 1MO-I3M. Also nitriles are formed, depending on the pH, the presence of specifier proteins, and Fe(II)ions. IAN is the I3M-derived nitrile. No organic thiocyanates and epithionitriles derived from indole glucosinolates are presently known. The glucobrassicin-derived indole-3-ylmethyl-ITC immediately converts in reaction with ascorbic acid to ascorbigen and upon reaction with water to I3C. Also inorganic thiocyanate ions (SCN^-) are produced. Substituted indole glucosinolates are broken down in a similar fashion as I3M (*60*). I3C and ascorbigen have been shown to have anti-cancer effects (*61*). *In vivo*, I3C is further processed to various oligomeric products such as 3,3'-diindolylmethane and its different derivatives, N-methoxyindole-3-carbinol, and indolo[3,2-b]carbazole, which are presumably responsible for the effects of I3C in mammals. Among others, I3C and its derivatives act anti-proliferative in tumor cells by leading to cell cycle arrest, apoptosis induction, chemosensitization of chemoresistant cancer cells, reduction of inflammation and metastasis. Moreover, effects of these compounds against deoxyribonucleic acid (DNA)-adduct formation, immunomodulation and tumorigenesis have been observed (*61-63*). On the other hand IAN, another breakdown product of I3C, can also induce cancer by a reacting with nitrite, which leads to the formation of carcinogens (*61*).

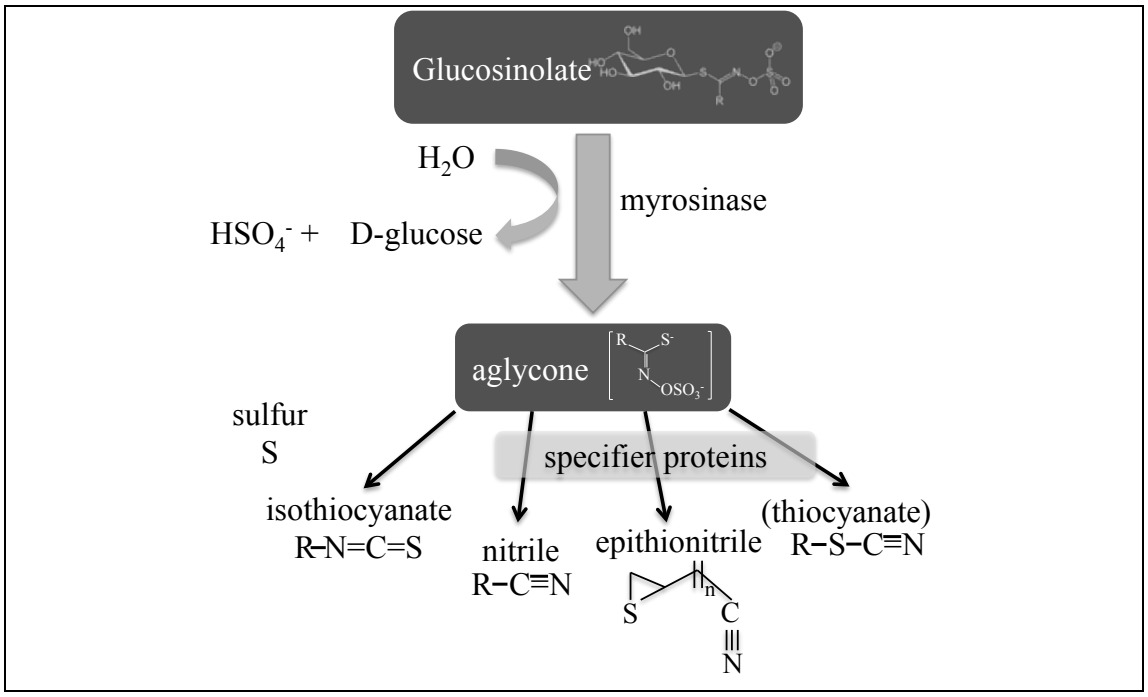


Fig. 3: Schematic and simplified diagram of glucosinolate hydrolysis by myrosinases. In the presence of specifier proteins, nitriles, epithionitriles and thiocyanates (not in Arabidopsis) are formed at the cost of ITC production. Adapted from (*27*).

1.1.3 Phenylpropanoids

Phenylpropanoids are compounds derived from the carbon backbone of the amino acid phenylalanine. They are produced via the phenylpropanoid pathway (see Fig. 4) (*4,64*). Phenylalanine is transformed into cinnamate, which in turn is transformed into *p*-coumaroyl Coenzyme A (CoA). *p*-Coumaroyl CoA is the precursor of a variety of compounds such as coumarins, flavonoids, lignin, etc. This diverse group of phenylpropanoids plays a role in the protection against ultra violet (UV)-light, response to mineral treatment, structural stability, pathogen defense, and reproduction (*4,64*). Examples of phenylpropanoids found in Arabidopsis are the coumarins such as scopolin and scopoletin, present in roots (*65*). A prominent example of a phenylpropanoid with chemoprotective properties is resveratrol. The stilbene resveratrol (3,4',5-trihydroxystilbene) is produced among others in grapes as a phytoalexin against fungal pathogens (*18,66*). Jang *et al.* (1997) could first demonstrate the chemoprotective effect of resveratrol (*67*). Numerous preclinical studies have been performed with this compound. The substance can both prevent the development and the progression of cancer (*68,69*). In a clinical study it could recently be shown that resveratrol can modulate enzyme systems involved in carcinogen activation and detoxification (*70*).

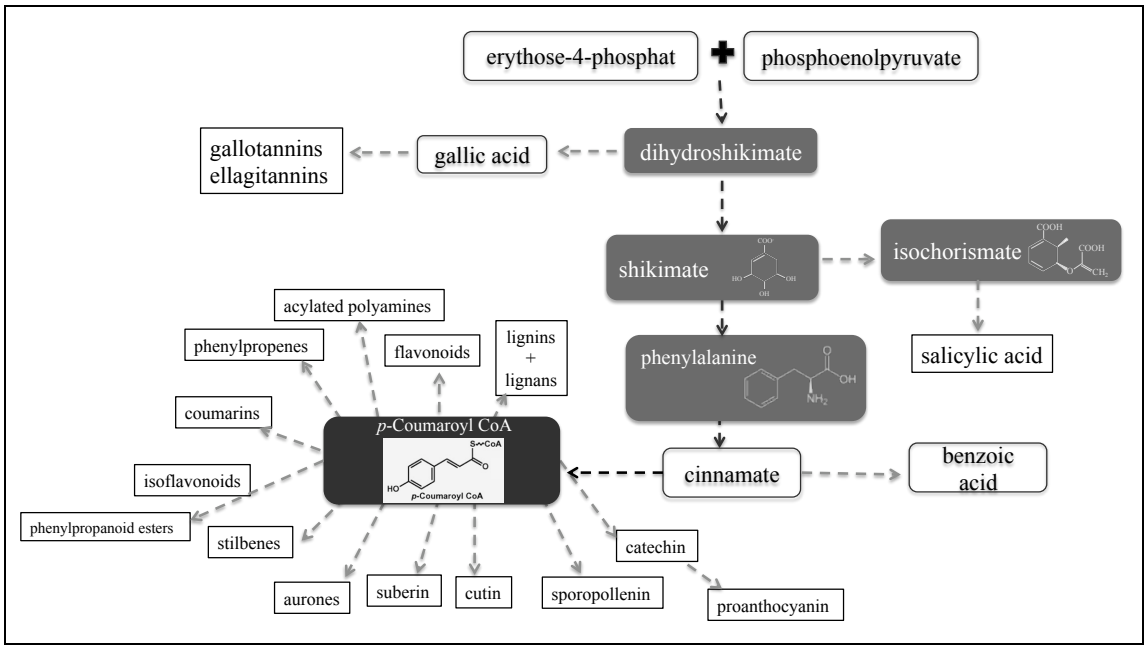


Fig. 4: Schematic diagram of the biosynthesis of phenylpropanoids. These compounds are derived from the amino acid phenylalanine. Related pathways are the synthesis of benzoid acid from cinnamate (common precursor) and the synthesis of salicylic acid (SA) from shikimate via isochorismate. Adapted from (*64*).

1.1.4 Benzenoids

Benzenoid compounds have various functions in Arabidopsis and are found as volatiles, in the stems of inflorescences, bound to the cell wall and in seeds (*4*). Salicylic acid (SA) is one example of an important signaling compound involved in plant defense reaction against different biotrophic and necrotrophic pathogens (*71*). Volatile benzenoid compounds such as methyl benzoate and methyl salicylate have been detected in floral scents. Another role of methyl salicylate is to attract predators of insects feeding on the plants (*72*). Benzoic acid can be found in form of glucosinolate esters in seeds, suggesting a role in storage of glucosinolates (*73*). An important human application of the benzenoid compound salicylate is in aspirin. SA has been shown to have anti-inflammatory effects in humans (*74*). Inflammation is a known causal agent of cancer (*75*). It was shown that sodium salicylate induces apoptosis and inhibits proliferation in human cancer cell lines (*76*). More than one pathway is predicted to exist for the production of benzoic acid. The benzenoids are probably derived from phenylpropanoids. Alternatively, it has been considered, that benzenoids are synthesized via the shikimate-derived isochorismate (Fig. 4). Most research in the past has been done on the phenylalanine pathway (*72*).

1.1.5 Flavonoids

Flavonoids are derived from the phenylpropanoid pathway (Fig. 4). They comprise the subclasses of flavanols, flavanones, anthocyanidins, isoflavones, flavones and flavonols (*77*). The most abundant flavonoids in Arabidopsis are the flavonols kaempferol glycosides, quercetin glycosides and an anthocyanin. This anthocyanin has a cyanidin core attached to four sugars, and the single residues *p*-coumaroyl, sinapoyl and malonyl (*4*). In plants, flavonoids have functions in auxin transport, defense against pathogens, allelopathy, ROS level modulation, flower coloring, and absorption of UV-light (*78*). Recently it was suggested that flavonoids might also have a function as an excess energy escape valve under stress conditions such as cold stress, acclimation and water deficit. In addition, the increased production of flavonoids are an advantage under these stress conditions (*79*). It has been shown that flavonoids can function as defense compounds against fungi, bacteria and insects in various plant species. Flavonoids also have signaling functions in symbiosis in many legumes and in wheat (*80-82*). Arabidopsis mutants lacking flavonoids are more sensitive to UV-B-light (*83*). Independent evidences support a role of flavonoids in auxin transport. The flavonoids apigenin, kaempferol and quercetin can inhibit auxin transport and could function as natural regulators of auxin transport and it has been shown that auxin transport is elevated in a mutant that lacks flavonoids (*transparent testa*) (*84-86*). Flavonoids are beneficial for humans due to their anti-oxidant, anti-proliferative, anti-tumor, anti-inflammatory and pro-apoptotic effects (*87*). For the phytochemical kaempferol it was recently demonstrated that it has inhibiting effects on the development of skin carcinogenesis, on the proliferation of human cancer cell lines and the inhibition of neoplastic cell transformation (*88-90*). Quercetin can also inhibit cell transformation and can among others react with ROS and in high concentrations initiate apoptosis (*91-94*).

1.1.6 Terpenes

A. thaliana produces a variety of terpene compounds. Terpenes represent the largest and most diverse class of metabolites in plants. The compounds are composed of C_5 isoprenoid units and are classified by their number (*4,95*). Two independent pathways exist for the biosynthesis of the general C_5-isoprenoid precursors isopentenyl

diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). These are the mevalonate pathway and the methylerythritol phosphate pathway (MEP). The mevalonate pathway can be found in most eukaryotes and in prokaryotes and consists of several enzymatic steps. The precursor compound of this pathway is acetyl CoA. Via the production of mevalonic acid, IPP and DMAPP are formed. The MEP pathway takes place in the plastids. Its precursors are pyruvate and glyceraldehyde-3-phosphate (GAP). An intermediate of this pathway is 1-deoxy-D-xylulose 5-phosphate (DXP). Products of the MEP pathway are among others chlorophylls, carotenoids and gibberellins. Fig. 5 illustrates the MEP and mevalonate pathways. Chlorophylls, carotenoids, and plastoquinone are terpenes with primary functions in photosynthesis. Other terpenes with primary function in the plant physiology are ubiquinone in respiration and cytokinins, sterols, brassinosteroids, gibberellins and abscisic acid in plant growth and development (*95*). Terpenes have been found to also function as phytoalexins against blast fungus in rice (*96*), in defense against herbivores and pathogens in conifers (*97*) and in tobacco (*98*). Volatile terpenes have been described that attract pollinators and can have functions in defense (*4, 99-102*). In addition it has been suggested that volatile isoprenoids have anti-oxidant properties in plants (*103*). Lycopene has anti-oxidant properties and has been shown in humans to have positive effects in the prevention of prostate, breast, cervical, ovarian, and liver cancer (*104*).

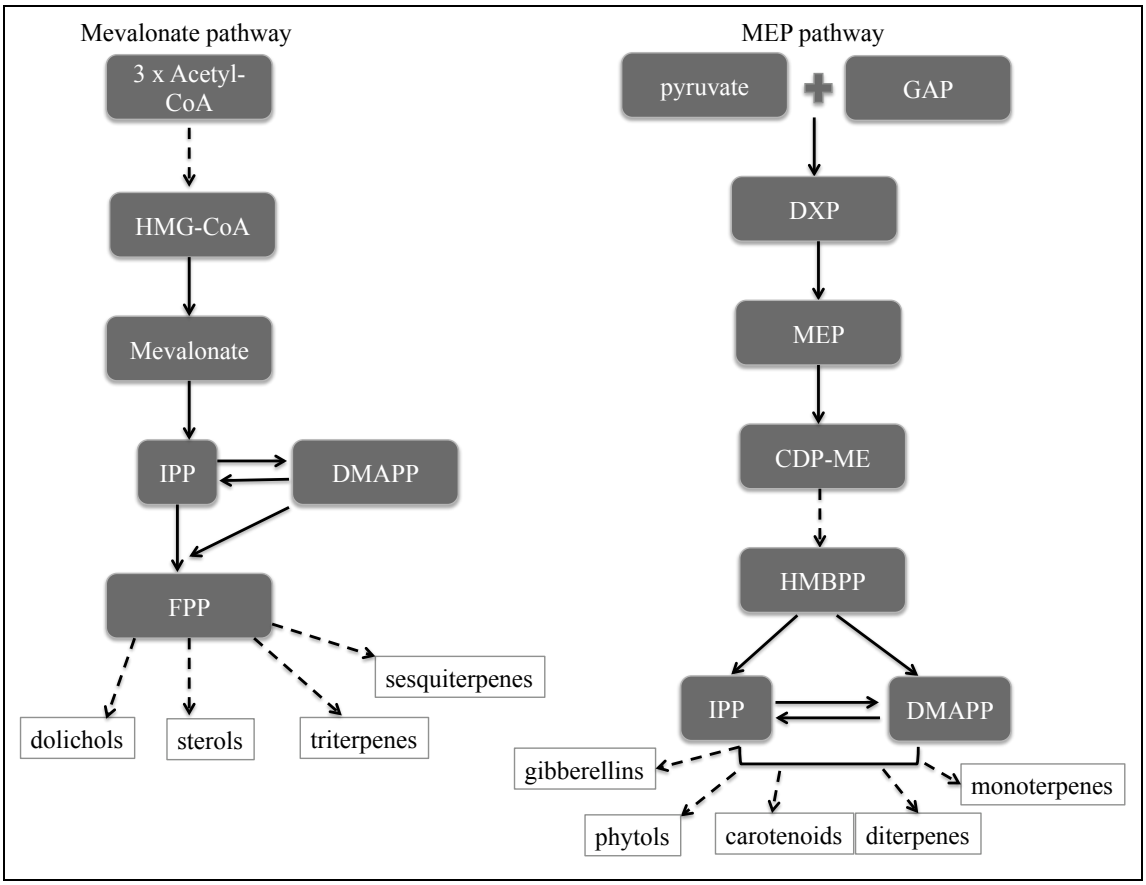


Fig. 5: Schematic diagram of the two terpene biosynthesis pathways. The two independent pathways are the mevalonate and the methylerythritol phosphate (MEP). The precursor of the mevalonate pathway is acetyl Co-A, while the MEP pathway is derived from pyruvate and glyceraldehyde-3-phosphate (GAP). Abbreviations: 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA), isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP), farnesyl diphosphate (FPP), 1-deoxy-D-xylulose 5-phosphate (DXP), 4-diphosphocytidyl-2-C-methyl-D-erythritol (CDP-ME), (*E*)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate (HMBPP). Adapted from (*95*).

1.1.7 Fatty acid derivatives

The derivatives of fatty acids are another class of compounds with multiple functions in plants. Substituted acids and alkanes are components of surface waxes and suberin. Green leaf volatiles, hydroperoxides and ketodienes are oxylipins. Oxylipins are formed from fatty acids by lipoxygenase action. (*4*) Jasmonic acid, an oxylipin, has important signaling function in plant defense in response to herbivory, wounding and disease (*105*). Arabidopsis infected with the caterpillar *Pieris rapae* attract the parasitic wasp *Cotesia rubecula* by emitting volatiles containing among others green leaf volatiles (*106*). Not only oxylipins play a role in plant defense, but also unsaturated fatty acids and very long chain fatty acids have been associated with plant defense by altering membrane lipid composition and signaling (*107,108*). Fatty acids

are synthesized in the plastids. The compounds are formed via chain elongation of acetyl-CoA with malonyl-CoA. Finally, a 16- or 18- carbon acyl chain is released by breakage of the thioesterbond between the carbon chain and the acyl carrier protein (*1*). There is also evidence that sphingolipids may play a role in cancer development in animals, since it could be shown that the digestion products of the sphingolipid sphingomyelin found in milk, ceramide and sphingosine, inhibit colon carcinogenesis (*109*). Fingrut and Flescher could demonstrate the anticancer effects of jasmonic acid and methyl jasmonate (*76*). Jasmonic acid leads to cell death in lymphoblastic leukemia cells and suppresses the proliferation of several different human cancer cell lines. Cell death of human cancer cells was also caused by methyl jasmonate. The compounds did not harm normal human cells (*76*).

1.2 Development of cancer and EpRE based chemoprevention

The development of cancer is generally divided into three different stages. The cells go through a process of initiation, followed by promotion and progression. Cancer is initiated upon uptake of or exposure to a carcinogenic agent, radiation, UV exposure, or inherited defect in genes required to prevent cancer such as DNA mismatch repair genes. The carcinogenic agent is distributed and transported to organs and tissues, where it is metabolically activated (*5*). The carcinogen covalently interacts with genomic DNA leading to the activation of oncogenes and/or inactivation of tumor suppressor genes, via DNA mutation. This phase of tumor development is a rapid and irreversible process. The next step of tumor development is the reversible promotion phase during which the initiated cells proliferate and accumulate. With the final progression stage of the cancer, the tumor develops invasive and metastatic potential (*5*).

In general xenobiotics are processed in the organism via phase I and phase II drug metabolizing enzymes. While phase I enzymes catalyze oxidation, reduction and hydrolysis of the compound, which can lead to the activation of a carcinogen, phase II enzymes catalyze glucuronidation, sulfation, acetylation, methylation, and conjugation. Cytochrome P450 monooxygenases are an important family of phase I activating enzymes (*110*). Oxidative stress is also a causal agent of carcinogenesis. ROS, such as hydrogen peroxide or superoxide anions are formed in the cells as a by-

product during aerobic metabolism. For instance, in inflamed tissues, immune cells produce ROS and reactive nitrogen species. Other ROS producers are microbial infection, UV-radiation, and toxic chemicals (*5,75*). Phase II detoxifying and anti-oxidant enzymes, among others glutathione *S*-transferase (GST), NAD(P)H quinone oxidoreductase 1 (QR) and heme oxygenase-1 (HO-1), protect cells against potential carcinogens. Anti-oxidant enzymes have a dual role. They inactivate ROS and eliminate and detoxify electrophiles. Therefore, they are not clearly distinguishable from phase II detoxification enzymes (*5*). The toxic compounds are converted to less reactive and toxic agents, via conjugation with endogenous substrates such as glutathione, glucuronide, or sulfate. This leads to an increase in their solubility and promotes their clearance from the organism (*8*).

Phytochemicals are divided into two categories, depending on their way of intervention during cancer development. Those belonging to the category of blocking agents induce the detoxification of carcinogens, inhibit their metabolic activation, reduce their synthesis, prevent the interaction of carcinogens with DNA and remove free radicals, therefore interfering with the initiation stage. The mechanism behind this way of function is the induction of the expression of phase II detoxifying and anti-oxidant enzymes. The second group consists of cancer suppressing agents that interfere with the promotion and progression of cancer. These compounds impact the cell proliferation, differentiation, senescence and/or apoptosis (*110*) (Fig. 6). Examples of blocking phytochemicals playing a dual role by inducing detoxifying/anti-oxidant enzymes and having anti-oxidant activity themselves are phenolic anti-oxidants such as curcumin, and caffeic acid or the stilbene resveratrol (*67,110*).

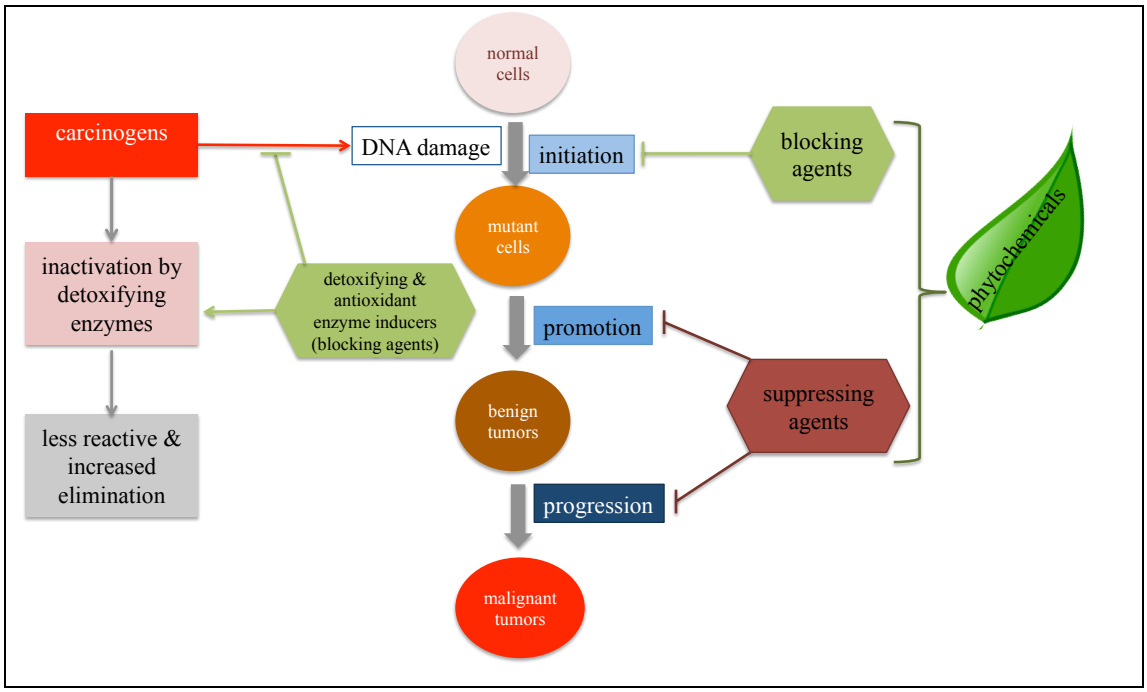


Fig. 6: Schematic diagram of the role of dietary phytochemicals as inducers of cytoprotective and anti-oxidant enzymes. The three stages of cancer development are initiation, promotion and progression. Phytochemicals can function as cancer blocking or cancer suppressing agents. Cancer blocking agents block the initiation of cancer by inhibiting carcinogens from damaging DNA, which would lead to mutations, or they induce detoxifying and anti-oxidant enzymes. This results in the inactivation of carcinogens, which are then less reactive and can be further processed by the organism. Adapted from (*110*).

The induction of phase II enzymes is mediated through a common mechanism, controlled by an electrophile response element (EpRE). EpRE is present in the upstream promotor region of genes encoding phase II enzymes. A member of the NF-E2 family of basic leucine zipper transcription factor, the nuclear factor-erythroid-2-related factor 2 (Nrf2) plays an essential role in the induction of phase II detoxifying/ anti-oxidant enzymes. Nrf2 is negatively regulated via the Kelch-like ECH-associated protein 1 (KEAP1), which is an actin-binding protein. KEAP1 prevents the nuclear accumulation of Nrf2 and enhances its proteasomal degradation. A treatment with chemoprotective substances delays the degradation of Nrf2. Two known mechanisms of induction of phase II enzymes are by Nrf2 stabilization and phosphorylation. The protein KEAP1 contains 27 cysteine residues, some of which are necessary for the interaction with the N-terminal Neh2 domain of Nrf2. A modification of the cysteine residues by chemoprotective substances can disrupt the KEAP1/Nrf2 complex. Activation of Nrf2 can also take place through phosphorylation by mitogen activated protein kinases, protein kinase C, phosphatidylinositol 3-kinase, and ribonucleic acid

(RNA)-dependent protein kinase-like endoplasmatic reticulum kinase. Some chemopreventive agents induce the expression of these enzymes (*8,111*) (Fig. 7). In addition, it could be observed that the phytochemicals allyl-ITC, I3C and parthenolide lead to an increase of Nrf2 protein levels but do not affect its degradation rate (*8*).

A number of phytochemicals have been shown to induce both CYPs and detoxifying enzymes. These compounds are known as bifunctional inducers. Monofunctional inducers on the other hand increase the expression of cytoprotective enzymes only. Two examples of monofunctional inducers are *tert*-butylhydroquinone and sulforaphane. Belonging to the bifunctional inducers are β-naphthoflavone and I3C. In general it is assumed that the effect of bifunctional inducers is stronger on the detoxifying enzyme induction in comparison to phase I enzyme induction (*110*).

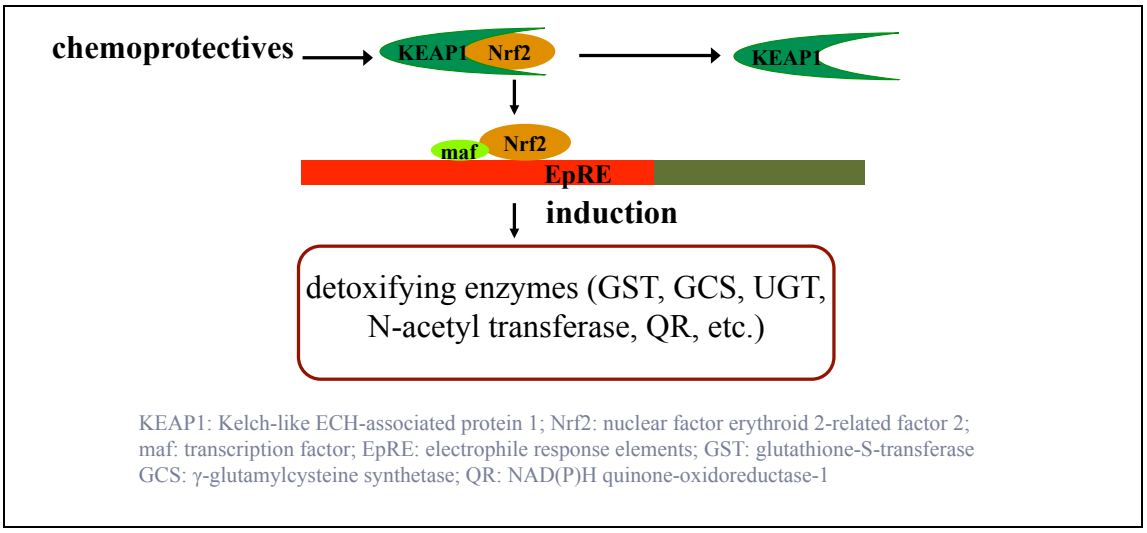


Fig. 7: Overview of the induction of detoxifying and anti-oxidant enzymes via chemoprotective compounds. Phase II and anti-oxidant enzymes have EpRE in their promotor region. The expression of these enzymes is regulated by the transcription factor Nrf2. Nrf2 is present in a complex with the transcription regulator KEAP1 in the cytosol. Chemoprotective compounds lead to the dissociation of the complex. Nrf2 can translocate into the nucleus and induce the expression of the detoxifying enzymes such as glutathione-*S* -transferase (GST) or NAD(P)H quinone oxidoreductase-1 (QR).

Assay systems have been developed that apply the EpRE promotor system to screen for chemoprotective agents in human and murine hepatoma cell culture (*112,113*). In these assay systems the reporter genes green fluorescent protein (GFP) and Luciferase are downstream of EpRE promotor elements in HepG2-GFP and Hepa1c1c7-Lux cells, respectively (Fig. 8). Upon treatment of the cells with phytochemicals the reporter genes are induced to different extent.

Fig. 8: Schematic diagram of the EpRE construct in human and murine reporter cells.

Different approaches have been used to screen for chemoprotective substances. Extracts of mutant and wild-type *A. thaliana* plants have been tested for QR inducing activity (*114*). Grubb *et al*. (2001) used the QR inducer potency of leaf extracts as an indicator for the total glucosinolate content in mutants (*115*). Wang *et al.* (2002) developed a simplified bioassay for the detection of glucosinolates by measuring the activity of QR in Hepa1c1c7 cells after treatment with excised leaf discs of *A. thaliana* (*116*). The measurement of QR activity has previously been applied in assay systems introduced by Prochaska *et al*. (*117,118*).

1.3 Secondary metabolite profiling in plants

Mass spectrometry (MS) is an analytical method to determine the molecular weight or the mass-to-charge ratio (*m/z*) of a compound (*119*). This highly sensitive method allows the elucidation of metabolite structures and it is frequently used for metabolite profiling (*120*). A metabolome is defined as a complete set of low molecular weight molecules present in a cell or organism. The molecules are part of metabolic reactions that are required for the growth, maintenance and function of an organism (*121,122*). The metabolome is influenced by gene expression, which is affected by the environment and mutations. Plants produce a large variety of compounds that differ in structure, physical and chemical properties, stability and quantity, thereby making metabolite profiling in plants very challenging. It is to date not possible to detect all compounds in plants with a single method (*11,120*). In order to analyze plant secondary metabolites, different analytical techniques have been developed, such as combinations of liquid chromatography (LC) with mass spectrometry (MS), gas chromatography (GC)- MS, and capillary electrophoresis (CE)- MS (*11,123*). The analysis of secondary metabolites begins with the sample collection followed by polar and semi-polar extraction using ethanol (EtOH) or water, or nonpolar extraction using chloroform (*11*). The pre-fractioning of compounds via GC, LC or CE reduces the target sample complexity before MS (*119*). After the fractioning, the compounds are analyzed using mass spectrometry. The MS consists of four different steps. The compounds are ionized with soft ionization techniques for example via electro spray

ionization (ESI) or matrix assisted laser desorption/ionization (MALDI), so that positively (positive mode) or negatively (negative mode) charged ions are produced. These are separated in the mass analyzer depending on their mass-to-charge (*m/z*) ratio. The ions can be detected either via their current or as image currents via orbital frequencies. Different mass analyzers have been developed such as time of flight (TOF), quadruple ion trap (QIT), linear ion trap (LIT), Fourier transform ion cyclotron resonance (FTICR), or quadrupole-time of flight (Q-TOF) (*119,122*). After the visualization of metabolite signals in a chromatogram, the peaks are compared with metabolite information from standard compounds. Upon assigning a specific peak to a metabolite, the detected metabolites of the analyzed samples are listed together with their intensity data (*11*). Not all metabolites can be identified in this way; therefore it is necessary to obtain information on the structure of a substance and to have the respective detailed compound libraries. The general MS procedure is summarized in Fig. 9.

Fig. 9: Schematic diagram of mass spectrometry (MS) used for the identification of metabolites. The metabolites are extracted from frozen plant material. After prefractionating of the target samples, via gas chromatography (GC), liquid chromatography (LC) or capillary electrophoresis (CE), they are subjected to MS analysis. MS consists of the ionization source (electro spray ionization (ESI) or matrix assisted laser desorption/ionization (MALDI)), the mass analyzer (time of flight (TOF), quadruple ion trap (QIT), linear ion trap (LIT), Fourier transform ion cyclotron resonance (FTICR)) and the detector. The collected data is processed with different software and analyzed statistically. Adapted from (*119*).

CE-MS can detect ionic metabolites for example sugar phosphates and nucleotides (*11*). GC-MS is mainly applied for the detection of primary metabolites (amino acids, organic acids, sugars, etc.), while LC-MS analysis can detect secondary metabolites, such as alkaloids, saponins, phenolic acids, phenylpropanoids, flavonoids, glucosinolates and polyamines (*11,124-128*). The latter method is performed with the high-resolution methods such as TOF, in order to identify new compounds. Ultra performance liquid chromatography (UPLC) -ESI-QTOF-MS analysis is applied in this work. Using this method, medium polar compounds can be detected in Arabidopsis; these include glucosinolates and their breakdown products, indole derivatives, phenylpropanoids, flavonoids/polyketides and fatty acid derivatives. It is not possible to detect terpenes with this protocol (*129*). The general workflow is summarized in Fig. 10. It begins with the growth of the plants and harvesting of the

plant material. After extraction of metabolites using aqueous EtOH, the samples are separated in the UPLC. The eluted compounds are detected via QTOF-MS. Different software is available for the processing of the LC-MS profiles. The data are further analyzed using various statistical methods.

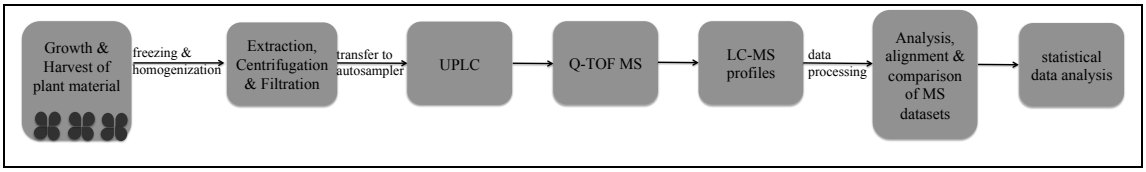


Fig. 10: Schematic workflow of an UPLC-ESI-QTOF-MS analysis for metabolite profiling of plant extracts. Adapted from (*124*).

1.4 Activation tagging and the TAMARA lines

Activation tagging is the use of enhancer elements that are located on a transfer-DNA (T-DNA) insertion tag. In this way, the expressed genes flanking the insertion site are up-regulated in a distance ranging from 0.4 to 8.2 kb. It is possible that even more than one gene, on both sides of the insertion, can be affected. Among the activation tagged plant populations are also knockout mutants that contain the insertion in the coding region of a gene. Typically, the *Cauliflower mosaic virus* (CaMV) promotor enhancer elements of the 35S gene are used (*130-132*). Walden and colleagues introduced this method for gene analysis (*133*). A number of activation tagged Arabidopsis populations have been generated, among others by Wilson *et al*. (1996), Tissier *et al*. (1999), Weigel *et al*. (2002), Marsch-Martinez *et al*. (2002), Nakazawa *et al*. (2003), Schneider *et al*. (2005) and Robinson *et al*. (2009) (*131,134-140*). Besides in Arabidopsis, activation tagging has been applied in several other plant species, namely poplar, rice, tomato, barley and petunia (*132*). The method of activation tagging has many advantages in comparison to classical loss of function mutants. A large part of the Arabidopsis genome consists of duplicated loci (*141*). Due to this circumstance, many of the genes can have redundant function and therefore the detection of many gene functions is not possible via loss of function mutations. The other extreme is that the loss of function can be lethal for the plant. It must also be considered that some phenotypes are only visible under certain conditions. The advantages of activation tagging are the dominant mutations, the viability of the plant versus a lethal loss of function mutation, it can overcome the effect of gene redundancy, positive selection screens can be performed under various stress conditions, and the opportunity to study metabolic pathways in plants is given

(*132*). Many compounds have not yet been detected in plants because they are present in concentrations below the sensitivity of current detection systems (*11,142*). Therefore, the activation tagged plants can allow the detection of compounds usually present at very low concentrations in wild-type plants. In Arabidopsis it has been observed that activation tagged genes can reflect the endogenous expression pattern of a gene in contrast to the ectopic expression caused by a full CaMV35S promotor (*136,143,144*). Visible phenotypes were observed in 0.1 – 1% of the lines of the populations (*131*). In this work the **t**ransposable element mediated **a**ctivation tagging **m**utagenesis in **Ara**bidopsis (TAMARA), a population of 9471 lines, is used (*135*). The TAMARA population contains approximately 6,000 independent random transposon insertions of the T-DNA containing a CaMV35S enhancer element tetramer. To ensure a non-specific and high frequent integration of the T-DNA into the Arabidopsis genome, the maize *En/Spm* transposon elements were used to generate the activation tagged population. *A. thaliana* Columbia (Col-*0*) plants were transformed with the TAMARA construct, containing the immobilized *Spm* transposase gene, under the control of the CaMV35S promotor, linked with the negative selectable marker SU1 and the mobile element containing the *bar* gene as a positive selection marker, linked to the 4 x 35S enhancer elements and the modified defective *Spm* element (*dSpm*-Act) (see Fig. 11) (*135*). The SU1 gene confers sensitivity to the pro-herbicide R7402 (DuPont), which leads to the production of the herbicide sulfonylurea. These plants are dwarfish, dark green and have reduced apical dominance. The *bar* gene confers resistance to the herbicide phosphinotricin (BASTA) (*132,145,146*). The plants were transformed with the TAMARA construct using *Agrobacterium tumefaciens* (*A. tumefaciens*). Transposition events were selected in the T_3-generation, via selection with both R7402 and BASTA.

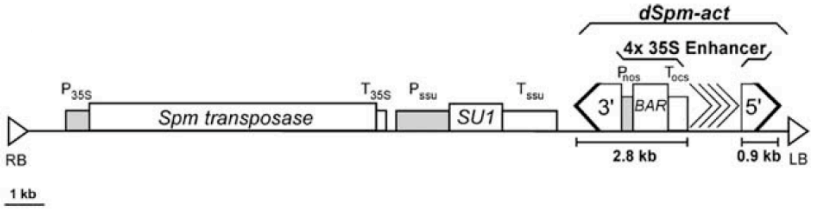


Fig. 11: Schematic diagram of the TAMARA construct. The TAMARA construct contains the *Spm* transposase gene, under the control of the CaMV35S promotor. The SU1 gene is linked to the transposase gene and is controlled by the *Rubisco small subunit* promotor. The T-DNA contains the defective *Spm* element, the CaMV35S enhancer tetramer, and the *bar* resistance gene. Picture taken from (*135*).

1.5 Aim of this work

Currently it is postulated that there are many unknown compounds present in Arabidopsis in concentrations below the sensitivity of current detection systems (*11,142*). The goal of this work is to identify novel chemoprotective compounds in Arabidopsis and to characterize genes involved in the synthesis and metabolism of these compounds. In a previous screen with HepG2-GFP cells, 470 independent mutant lines showing a high EpRE induction potential were selected. After re-cultivation and extract preparation of these candidate lines, they will be screened using a more sensitive assay system with Hepa1c1c7-Lux cells. From the expected positive mutants, the T-DNA insertion sites will be mapped and the expression of the genes flanking the insertion sites will be determined in relation to the expression in wild-type plants. Selected mutants are then planned to undergo untargeted metabolite profiling using mass spectrometry in collaboration with Prof. D. Scheel, Dr. Stephan Schmidt, and Dr. Christoph Böttcher (IPB, Halle). Further analyses of mutants selected on the basis of available metabolite profiles or the putative gene functions of the affected genes are planned.

2 Materials & Methods

2.1 Materials

2.1.1 Instruments, equipment and disposable materials

The manufacturer of the instruments and equipment used during the laboratory work are specified in the corresponding method parts. Disposable material was supplied by VWR (Darmstadt, D).

2.1.2 Chemicals and kits

The analytical chemicals were obtained from the companies Biomol (Hamburg, D), Difco (Hamburg, D), Duchefa (Haarleem, NL), Fluka (Buchs, CH), Merck AG (Darmstadt, D), Roth (Karlsruhe, D), Sigma-Aldrich (Deisenhofen, D) and Serva (Heidelberg, D). The applied kits and ready to use solutions are listed in Table 1. Media and solutions for animal cell culture are listed in Table 2.

Table 1: List of kits

Name	Supplier
BigDye ® Terminator V3.1 cycle sequencing kit	Applied Biosystems, (Foster city, USA)
E.Z.N.A.® gel purification kit	Omega Bio-Tek, Inc. (Norcross, GA, USA)
Quantum prep ® plasmid miniprep kit	BioRad, (München, D)
BCA™ protein assay kit	Pierce, (Rockford, USA)
MinElute Gel Extraction Kit	Qiagen, (Hilden, D)
QIAquick Gel Extraction Kit	Qiagen, (Hilden, D)
Superscript II Reverse Transcriptase Kit	Invitrogen, Life Technologies (Karlsruhe, D)
SYBR GREEN PCR mix	Applied Biosystems, (Darmstadt, D)
SensiMix™ SYBR® kit	Bioline Reagents Ltd., (London, GB)
Brilliant SYBR® Green QPCR Core Reagen kit	Agilent Technologies (Böblingen, D)

Table 2: Ready-to-use media and solutions for animal cell culture

Name	Supplier
DPBS	Gibco®, invitrogen
Geneticin G418	Gibco®,invitrogen
L-Glutamine	Gibco®,invitrogen
Pen-Strep	Gibco®,invitrogen
Recovery freezing media	Gibco®, Invitrogen
RPMI 1640	Gibco®, invitrogen
α-MEM	Gibco®, invitrogen
accutase	PAA$_{\text{The cell culture company}}$
trypsin-EDTA	Invitrogen

2.1.3 Enzymes

The enzymes for molecular biological work were obtained from the companies MBI Fermentas GmbH (St.Leon-Rot, D), Roche Molecular Biochemicals (Mannheim, D), Qiagen GmbH (Hilden, D) and New England Biolabs (Frankfurt/Main, D). The enzymes and suppliers are specified in the corresponding method chapters.

2.1.4 Vectors

The vectors used for cloning and protein expression are listed in Table 3.

Table 3: List of vectors

Vector	Application and reference
pENTR-1A	Gateway compatible entry clone via classical cloning; kanamycin resistance gene (Invitrogen, Karlsruhe, D).
pENTR-4	Gateway compatible entry clone via classical cloning; kanamycin resistance gene (Invitrogen, Karlsruhe, D).
pDONR207	Gateway donor vector via BP-reaction; gentamycin resistance gene (Invitrogen, Karlsruhe, D).
pGWB2	Expression vector; 35S CaMV promotor for overexpression of a target gene *in planta*, kanamycin and hygromycin resistance genes for selection in plants and in bacteria.
pGWB3i	Expression vector with C-terminal GUS gene for promotor analysis *in planta*; modified version of pGWB3 (*147*) with a plant intron in the GUS gene to inhibit bacterial expression (*148*). Kanamycin and hygromycin resistance genes for selection in plants and in bacteria.
pGWB5	Expression vector with 35S CaMV promotor and production of C-terminal GFP fusion protein. Kanamycin and hygromycin resistance genes for selection in plants and in bacteria (*147*).
pGWB6	Expression vector with 35S CaMV promotor and production of N-terminal GFP fusion protein. Kanamycin and hygromycin resistance genes for selection in plants and in bacteria (*147*).
pAMPAT	Gateway compatible expression vector with 35S CaMV promotor. Ampicillin/Carbenicillin resistance gene for selection in bacteria. Bar gene for selection with BASTA® *in planta*. Derivative of pAMPAT-MCS binary vector (Lipka V, Rademacher T, Panstruga R, 2003).

2.1.5 Antibiotics

Table 4 lists the antibiotics used for the selection of transformed *Escherichia coli* (*E. coli*) and *A. tumefaciens*, as well as for the selection of transgenic *A. thaliana* plants.

Table 4: List of antibiotics and concentrations

Antibiotic	Concentration	Solvent	Supplier
amphotericin	5 µg/mL (*A. thaliana*)	H_2O	Sigma
carbenicillin	100 µg/mL	Tris HCl pH 8.0	Serva, Duchefa
chloramphenicol	10 µg/mL (*E. coli* DB3.1) 75 µg/mL (*A. tumefaciens*)	EtOH	Duchefa, Roth, Sigma
gentamycin	25 µg/mL (*A. tumefaciens* SV-0) 10 µg/mL (*E. coli*)	H_2O	AppliChem
hygromycin	50 µg/mL 50 µg/mL (*A. thaliana*)	H_2O	Duchefa
kanamycin	50 µg/mL 50 µg/mL (*A. thaliana*)	H_2O	Duchefa
phosphinothricin BASTA	10 µg/mL (*A. thaliana*) 250 mg/L	H_2O	Dr. Ehrenstorfer GmbH Bayer
rifampicin	20 µg/mL (*A. tumefaciens* SV-0) 50 µg/mL (*A. tumefaciens* GV3101 pMP90, GV3101 pMP90RK) 150 µg/mL (*A. tumefaciens* P19)	DMSO	Duchefa
ticarcillin	160 µg/mL (*A. thaliana*)	H_2O	Duchefa

2.1.6 Oligonucleotides

The oligonucleotides (primers) were supplied by Metabion GmbH (Martinsried, D) and Sigma-Aldrich® (Deisenhofen, D). For cloning purposes the primers were designed using Vector NTI® software (Invitrogen, Karlsruhe, D). The oligonucleotides for Real time PCR (polymerase chain reaction) were designed using the software primer Express® (Applied Biosystems, Darmstadt, D). The complete list of oligonucleotides can be found in the appendix.

2.1.7 Organisms

2.1.7.1 Plants

The leaves of the tobacco plants *Nicotiana benthamiana* (*N. benthamiana*) were used for subcellular localization studies via transient expression of GFP fusion proteins. *A. thaliana* L-Heyn cv. Columbia-*0* (Col-*0*) (Nottingham Arabidopsis Stock Center (NASC)-Nr. N1093) functioned as a background for the overexpression of target genes and the study of loss of function mutations. Table 5 is a list of *A. thaliana* mutants used in this work.

Table 5: List of *A. thaliana* mutants

Mutant	Species, ecotype	Origin
TAMARA lines	*A. thaliana,* Col-*0*	Schneider *et al.* (2005) (*135*)
WiscDsLox461-464F2 *pyk10-1*	*A. thaliana,* Col-*0*	Nottingham Arabidopsis Stock Center
myb11/myb12/myb111	*A. thaliana,* Col-*0*	Stracke *et al.* (2007) (*149*)
MIM858	*A. thaliana,* Col-*0*	Todesco *et al.* (2010) (*150*)
hig1-1	*A. thaliana,* Col-*0*	Gigolashvili *et al*. (2007) (*151*)

2.1.7.2 Bacterial strains

Table 6: Bacterial strains

Strain	Genotype, reference
DB3.1 (*E. coli*) cultivation of plasmid containing ccdB	F^- *gyr*A462 *end*A1 Δ(sr1-*rec*A) *mcr*B *mrr* *hsd*S20(r_B-, m_B-) *sup*E44 *ara*-14 *gal*K2 *lac*Y1 *pro*A2 *rps*L20(Sm^R) *xyl*-5 λ- l*eu mtl*1 (Invitrogen, Karlsruhe, D)
DH5α (*E. coli*) cultivation of plasmid DNA	F^- Φ 80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *deoR* *recA*1 *endA*1*hsdR*17(rK^- mK^+) *phoA supE*44 $λ^-$ *thi*-1 *gyrA*96 *relA*1 (*152*)
GV3101 (*A. tumefaciens*) stable plant transformation	pMP90, Rif^R, $Gent^R$ (*153*)
GV3101 (*A. tumefaciens*) stable plant transformation with vector pAMPAT	pMP90RK, Rif^R, $Gent^R$, Kan^R
P19 (*A. tumefaciens*) anti-silencing strain	C58C1-pCH32-35S:p19, Rif_R, Kan_R (*154*)
SV0 (*A. tumefaciens*) transient expression in leaves and cultured cells	LBA4404.pBBR1MCS *vir*GN54D Rif^R, $Chlor^R$ (*155*)
TOP10 (*E. coli*) cultivation of plasmid DNA	F^-*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ 80d*lacZ*ΔM15 Δ*lac*X74 *rec*A1*ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G (Invitrogen, Karlsruhe, D)

2.1.7.3 Human and murine cell culture

Prof. Dr. William E. Fahl (University of Wisconsin, USA) provided HepG2 cells containing the EpRE-TK-GFP construct. The human liver epithelial cells are derived from a hepatocellular carcinoma of a 15-year-old Caucasian boy. The Hepa1c1c7 cells were provided by the American Type culture collection (Manassas, VA, USA). The *Mus musculus* derived hepatoma cells are liver epithelial cells. The Hepa1c1c7-Lux (EpRE(mGST-Ya)-LUX) cells were kindly provided by Dr. Aarts (Wageningen University, Netherlands). The cells contain the EpRE-LuX construct.

2.1.8 Computer software

Table 7: Computer software

Software	Manufacturer
Edit view	Applied Biosystems
Fluorescence microscope, binocular	Intas GDS
Gel documentation	Intas GDS
i-control 1.4	Tecan Group Ltd. Infinite 200
Magellan6	Tecan Group Ltd. Infinite 200
primer express	Applied Biosystems
SDS software	Applied Biosystems
Vector NTI	Invitrogen
ZEN lite 2009	Zeiss

2.2 Microbiological methods

2.2.1 Preparation of media

Lysogeny broth medium (LB-medium) from Sambrook & Russel (2001) (*156*)

5 g yeast extract
10 g peptone
10 g NaCl
(12 g agarose)

ad 1000 mL with $H_2O_{millipore}$

autoclave

Yeast extract and beef medium (YEB-medium)

5 g bacto peptone
1 g yeast extract
5 g beef extract
5 g saccharose
0.5 g $MgSO_4 \times 7H_2O$
(15 g agarose)

ad 1000 mL with $H_2O_{millipore}$

autoclave

Super optimal broth medium (SOB-medium)

0.5%	yeast extract
2%	tryptone
10 mM	NaCl
2.5 mM	KCl
10 mM	$MgCl_2$
10 mM	$MgSO_4$
$H_2O_{millipore}$	

For the preparation of SOB-medium yeast extract, tryptone, NaCl and KCl were suspended in $H_2O_{millipore}$. After autoclaving, sterile filtered $MgCl_2$ and $MgSO_4$ were added to the medium.

Super optimal broth with Catabolite Repression (SOC-medium)

SOC-medium consists of SOB-medium with 20 mM sterile filtered glucose.

2.2.2 Cultivation of bacteria

2.2.2.1 Cultivation of *E. coli* cells on LB-plates

E. coli cells were evenly spread on LB-plates containing the appropriate antibiotics for selection. The bacteria were incubated for 12-24 h at 37 °C.

2.2.2.2 Cultivation of *E. coli* cells in liquid medium

4 mL of liquid LB medium, containing the appropriate antibiotics, was inoculated with *E. coli* cells from one colony. The bacteria were cultivated at 37 °C on a shaking incubator 220 rpm for 24 h (Multitron, Infors GmbH, Einsbach, D).

2.2.2.3 Cultivation of *A. tumefaciens* on YEB-plates

A. tumefaciens cells were spread on YEB-plates containing the appropriate antibiotics for selection. The bacteria were incubated, depending on the bacterial strain, for 2 to 5 days at 28 °C.

2.2.2.4 Cultivation of *A. tumefaciens* in liquid medium

Single colonies of *A. tumefaciens* were used to inoculate liquid YEB-medium, containing the appropriate antibiotics for selecting the bacterial strain and the plasmid. Depending on the strain, the agrobacteria could be inoculated in the final needed volume of YEB medium (LBA4404, GV3101) or first in 1 mL of YEB (GV3101

pMP90RK), which was then used to inoculate larger volumes (50 mL) of YEB medium. The bacteria were grown at 28 °C on a shaking incubator at 220 rpm (Multitron, Infors GmbH, Einsbach, D).

2.2.2.5 Long-term storage of bacteria

For long-term storage of transgenic bacteria, these were grown in the respective liquid medium with antibiotics and mixed in equal volumes (0.5 - 1 mL each) with 60% glycerol. The mixture was immediately frozen in liquid nitrogen and stored at -80 °C.

2.2.3 Production of competent bacteria

For the production of competent bacteria, different protocols were used, depending on the bacteria and method of transformation.

2.2.3.1 Chemically competent *E. coli* cells (DH5α, TOP10)

100 - 500 mL Psi broth was inoculated with 200 to 1000 µL of pre-cultured *E. coli* cells. The bacteria were cultured on a shaking incubator at 37 °C and 220 rpm (Multitron, Infors GmbH, Einsbach, D), until they reached an OD_{600} of 0.6 – 0.7. The cells were incubated on ice for 15 min and then centrifuged at 4000 x *g* for 5 min (4 °C) (Eppendorf centrifuge, 5810 R). The bacterial pellet was resuspended in 0.4 volume of TfBI. After incubation on ice for 15 min, the cells were centrifuged (4000 x *g*, 5 min, 4 °C) and the pellet resuspended in 0.04 volumes of TfbII. Aliquots of 100 µL each, were prepared, after the cells had incubated on ice for 15 min, and were immediately frozen in liquid nitrogen and stored at -80 °C.

Psi broth:	5 g/L bacto yeast extract 20 g/L bacto tryptone 5 g/L $MgSO_4$ (pH 7.6 NaOH) autoclave	TfbI :	30 mM potassium acetate (CH_3CO_2K) 100 mM RbCl 10 mM CaCl 50 mM $MnCl_2$ 15% v/v glycerol (pH 5.8 acetic acid) filter sterilize	TfbII:	10 mM MOPS 75 mM CaCl 10 mM RbCl 15% (v/v) glycerol (pH 6.5 NaOH) filter sterilize

2.2.3.2 Chemically competent *A. tumefaciens*

A. tumefaciens were grown on a shaking incubator (220 rpm, Multitron, Infors GmbH, Einsbach, D) for 12-16 h in 5 mL YEB, containing antibiotics The bacterial pre-culture was used to inoculate 500 mL of YEB medium, which were grown for 4 - 6 h until they reached an OD_{600} 0.5-0.6. The bacterial culture was centrifuged at 7000 x *g* at 4 °C for 5 min. The bacterial pellet was resuspended in 25 mL (4 °C) 0.15 M NaCl and incubated for 15 min. The cells were centrifuged for 5 min at 7000 x *g* and 4 °C. The pellet was resuspended in a total volume of 5 mL (4 °C) 20 mM $CaCl_2$. Aliquots of 50-100 µL were immediately frozen in liquid nitrogen. The competent cells were stored at -80 °C.

2.2.3.3 Electro competent *A. tumefaciens*

Two flasks, each containing 5 mL of MGL-medium, were inoculated with agrobacteria. The agrobacteria were cultivated on a shaking incubator at 28 °C for 24 - 48 h at 220 rpm (Multitron, Infors GmbH, Einsbach, D). The pre-culture was diluted in 100 mL MGL-medium (OD_{600}: 0.04- 0.08). The bacteria were cultivated until they reached an OD_{600} of 0.5. After a centrifugation step for 15 min at 4,000 x *g* the pellet was resuspended in 40 mL 1 mM HEPES buffer (pH 7.0). All steps were performed on ice. After another centrifugation step, the bacterial pellet was resuspended in 40 mL 1 mM HEPES buffer (pH 7.0)/ 10% glycerol. After the next centrifugation the pellet was resuspended in 2 mL 1 mM HEPES pH 7.0/ 10% glycerol. The bacteria were evenly distributed between two reaction tubes. The two pellets were each resuspended in 200 µL 1 mM HEPES pH 7.0/ 10% glycerol. 50 µL aliquots were immediately frozen in liquid nitrogen and stored at -80 °C.

MGL-medium:	1 mM HEPES/ KOH pH 7.0	1 mM HEPES/ KOH pH 7.0/ 10% glycerol
5 g/L bacto-tryptone 2.5 g/L yeast extract 5 g/L NaCl 5 g/L mannitol 1.16 g/L Na-glutamate 0.25 g/L KH_2PO_4 0.1 g/L $MgSO_4$ x $7H_2O$ 1 mg/L bithin (stock: 10 mg/mL)		

2.2.4 Transformation of competent bacteria

2.2.4.1 Transformation of chemically competent *E. coli* cells

Competent *E. coli* cells were thawed on ice. For transformation 50-200 ng of plasmid DNA, BP, LR or ligation reaction (see chapter 2.4.3.1) was added to the cells. The samples were incubated on ice for 30 min, before being subjected to a 90 s heat shock in a 42 °C water bath. The samples were cooled on ice and 1 mL of LB medium or SOC medium was added to the cells. After shaking the bacteria at 37 °C (220 rpm, Multitron, Infors GmbH, Einsbach, D) for one hour, 200 – 1000 µL of cells were plated onto selective LB-plates. These were incubated at 37 °C for 12 -14 h.

2.2.4.2 Transformation of chemically competent *A. tumefaciens* cells

50-200 ng of plasmid DNA was added to the thawed competent *A. tumefaciens* cells. The mixture of DNA and agrobacteria was frozen in liquid nitrogen and then thawed again and subjected to a 2 min heat shock at 42 °C. 800 µL of YEB was added to each sample. The bacteria were incubated at 28 °C for 2 h and shaken with 220 rpm (Multitron, Infors GmbH, Einsbach, D). Different amounts of agrobacteria were plated on several YEB selection plates.

2.2.4.3 Electroporation of competent *A. tumefaciens* cells

200 ng of plasmid DNA was pipetted to 50 µL of thawed competent *A. tumefaciens* and incubated on ice for 2 min. The agrobacteria were transferred into a pre-cooled 2 mm electroporation cuvette (PEQlab, Erlangen, Germany) and subjected to 2 s electroporation using the electroporator Genepulser II (Bio-Rad, München, D) with the following parameters: 25 µF; 400 Ω; 2.5 kV; 8 – 9 ms

Immediately after electroporation 1 mL of YEB was added to the cells and they were incubated on the shaker for 2 h at 28 °C and 220 rpm (Multitron, Infors GmbH, Einsbach, D). The bacteria were plated on several selective YEB plates in different amounts.

2.3 Plant cultivation

2.3.1 Preparation of ½ MS-medium

10 g	saccharose
2.3 g	Murashige & Skoog Medium with modified vitamins (Duchefa, Haarleem, NL)
8 g	plant agar/gelrite
ad 1000 mL	with $H_2O_{millipore}$

pH 5.6 with KOH
(autoclave)

2.3.2 Seed sterilization

For sterilization, the seeds were transferred into a 1.5 mL reaction tube and incubated for 2 min in 70% EtOH. The seeds were mixed well with the EtOH. After the seeds had settled on the bottom of the reaction tube, the 70% EtOH was removed with a pipette. This step was repeated with 96% EtOH. The seeds were dried under the sterile hood and transferred onto sterile plates with ½ MS-medium and selective antibiotics.

2.3.3 Cultivation of *A. thaliana* on soil

A. thaliana seeds were germinated in plastic pots with 6 cm diameter (Pöppelmann, Lohne). The substrate was a mixture of Einheitserde® type VM and Perligran® (ca. 100 – 150 g/ 70 L). After the seeds were sown, they were covered with a plastic lid, to maintain a high humidity, and stratified for 2 to 3 days in the dark at 4 °C. The pots were then transferred into a growth chamber/cabinet or greenhouse. After approximately 2 weeks, the seedlings were separated and planted into single 6 cm pots or into multi-plates for 77 plants. Transgenic *A. thaliana* lines containing the *bar* gene were selected with BASTA®. The BASTA® solution consisted of 250 mg/L BASTA and 0.1% Tween-20. The plants were sprayed with the solution when they had developed four leaves.

components: Einheitserde® type VM:	natural clay white peat sod peat	pH 5.8 KCl (g/L): 1.0 N (mg/L): 100 P_2O_5 (mg/L): 180 K_2O (mg/L): 180 S mg/L: 200 Mg (mg/L): 500

2.3.4 Growth conditions

2.3.4.1 Green house

In the green house the plants were cultivated under long day conditions (with 16 h light and 8 h of darkness) and a relative humidity of 40%. The temperature during the dark was 18 °C, in the light 21 °C. The photon flow density of the photosynthetic active radiation varied due to outside conditions.

2.3.4.2 Plant growth cabinet or chamber

To reduce the variation between the plants and to increase the reproducibility of the results the plants were cultivated in growth cabinets or chambers. The conditions were either long day or short day, which 16 and 8 h of light, respectively. During the light period the temperature was 21 °C, it was reduced to 18 °C during the dark. The air humidity was 40%. The light intensity was 120 µmol $m^{-2}s^{-1}$. Under high light conditions the light intensity was 280 µmol $m^{-2}s^{-1}$.

2.4 Molecular biology

2.4.1 Extraction of nucleic acid

2.4.1.1 DNA extraction

Genomic DNA was extracted from *A. thaliana* leaves according to the protocol of Edwards *et al.* (1991) (*157*). The leaf material was immediately frozen in liquid nitrogen after harvest. The leaf tissue was homogenized for 1 min at 20 Hz using the tissue lyser (Qiagen, Hilden, D). 400 µL of extraction buffer was added and the samples were vortexed for 5 s (VortexGenie 2, Scientfic Industries, Inc, Bohemia, NY, USA). After 5 min of centrifugation (Minispin, Eppendorf AG, Hamburg, D) at

12,000 x *g*, 300 µL of supernatant was transferred into a new tube and mixed with 300 µL 2-propanol. The samples were incubated at room temperature for 2 min and then centrifuged for 5 min at 12,000 x *g*. The DNA pellet was vacuum dried using a concentrator (concentrator 5301, Eppendorf AG, Hamburg, D) and resuspended in 100 µL of TE-buffer. The genomic DNA was stored at -20 °C.

DNA extraction buffer:	200 mM Tris HCl, pH 7.5 250 mM NaCl 25 mM EDTA 0.5% SDS	1 x TE buffer:	10 mM Tris HCl, pH 7.4 1 mM EDTA, pH 8.0

2.4.1.2 Genomic DNA extraction from inflorescences

For a high yield of pure genomic DNA, 4 – 6 inflorescences and/ or leaf material was collected and immediately frozen in liquid nitrogen. Each sample was homogenized in 250 µL of extraction buffer using a tissue lyser (Qiagen, Hilden, D). After addition of further 250 µL extraction buffer, the samples were mixed with 400 µL of PCI. The samples were centrifuged for 15 min at 1,300 x *g* (Eppendorf centrifuge, 5417R). The supernatant was mixed with 800 µL 2-propanol and incubated for 10 min at room temperature. The precipitate was pelleted at 1,300 x *g*, for 10 min. The DNA pellet was washed twice in 70% EtOH and centrifuged after each washing step for 10 min at 500 x *g*. Succeeding drying of the DNA, the pellet was resuspended in 40 µL of TE buffer (pH 8.0). The DNA was stored at -20 °C. The quality of the DNA was checked by separation 1 µL of genomic DNA in a 1% agarose gel.

The DNA extraction buffer was prepared shortly before DNA extraction by adding the respective amounts of urea (from 12 M urea stock solution) and phenol to the 2 x DNA extraction buffer (stock solution).

DNA extraction buffer:	0.3 M NaCl 50 mM Tris pH 7.5 20 mM EDTA 2% (w/v) sarcosyl 0.5% (w/v) SDS 4.8 M urea 5% (v/v) phenol	2 x Buffer for DNA extraction:	0.6 M NaCl 100 mM Tris pH 7.5 40 mM EDTA 4% (w/v) sarcosyl 1% (w/v) SDS

phenol/chloroform/isoamylalcohol (25:24:1) (PCI)

70% EtOH

1 x TE: 10 mM Tris pH 7.5; 1 mM EDTA pH 8.0

2.4.1.3 RNA extraction

2.4.1.3.1.1 RNA extraction using TRIzol (Invitrogen) or TRIsure (Bioline) reagent

50 – 100 mg of plant material was harvested and immediately frozen in liquid nitrogen. The tissue was homogenized and 800 µL of TRIzol was added to the samples. The samples were mixed well and then centrifuged at 18,000 x *g* for 10 min (4 °C) (Eppendorf centrifuge, 5417). The supernatant was transferred into a 2 mL reaction tube and the samples were incubated at room temperature for 5 min. 160 µL of chloroform was added to the samples, after mixing them thoroughly, they were incubated at room temperature for 2 – 3 min. The samples were centrifuged at 12,000 x *g* for 15 min, at 4 °C. The aqueous phase was transferred with a pipette to a new 1.5 mL reaction tube. The RNA was precipitated by the addition of 0.3 mL 2-propanol. After 10 min incubation at room temperature, the samples were centrifuged for 10 min (4 °C and 12,000 x *g*). The RNA was washed with 600 µL of 75% EtOH (H_2O_{DEPC}). After centrifuging the samples for 5 min, at 4 °C and 7,500 x *g*, the RNA pellet was dried and resuspended in 30 µL ultra pure RNase free water. The RNA was incubated for 10 min at 60 °C.

2.4.1.3.1.2 RNA extraction using Z6-buffer

After harvesting of the plant material (50 – 100 mg), it was immediately frozen in liquid nitrogen to reduce RNase activity. The plant material was homogenized using the tissue lyser (Qiagen, Hilden, D). The powdery material was immediately frozen again in liquid nitrogen. 800 µL of Z6-buffer was added to the samples, after they were placed on ice. After mixing the samples with the Z6-buffer, they were centrifuged for 15 min at 17,950 x *g* and 4 °C (Eppendorf centrifuge, 5417). The supernatant was pipetted into a new 2 mL reaction tube. 500 µL of freshly prepared phenol/chloroform/isoamylalcohol (25/24/1) was added to each sample. After mixing the samples thoroughly, they were incubated for 2 min at room temperature and then centrifuged for 15 min at 12,000 x *g* and 4 °C. The aqueous phase was transferred into a 1.5 mL reaction tube and mixed with 1/20 volume of 1 M acetic acid and 0.7 volume of 96% EtOH. The RNA was precipitated at -20 °C for 1 h. The RNA was pelleted at 18,000 x *g* for 15 min (4 °C) and washed in 500 µL 3 M sodium acetate and in 500 µL 70% EtOH, each step followed by centrifugation (15 min, 18000 x *g*, 4

°C). The pellet was dried using the vacuum concentrator (Eppendorf concentrator 5301) and resuspended in 25 – 40 µL of $H_2O_{ultrapure}$. The RNA was incubated on a shaker at 65 °C until the RNA was completely resuspended. DNA contaminations in the samples were digested as described in chapter 2.4.3.6.

Z6-buffer:	8 M guanidinium-hydrochloride 20 mM MES 20 mM EDTA, (pH 7.0) 0.7% (v/v) β-mercaptoethanol (added shortly before use) H_2O_{DEPC}	1 M acetic acid: (H_2O_{DEPC})	3 M sodium acetate: (H_2O_{DEPC}) (pH 5.0 acetic acid)	75% EtOH (H_2O_{DEPC})

0.01% DEPC H_2O:

100 µL DEPC in 1000 mL H_2O

The H_2O was stirred using a magnetic stirrer over night and then autoclaved twice.

2.4.2 Plasmid mini preparation from *E. coli*

4 mL of *E. coli* cells, grown in liquid LB-medium for 12 – 16 h, were centrifuged in 2 mL reaction tubes (1 min, 12,000 x *g*) (Minispin, Eppendorf AG, Hamburg, D). The supernatant was removed and the bacterial pellet was resuspended in 200 µL of solution 1. 200 µL of solution 2 was added to the samples. After inverting the tubes several times, 200 µL of solution 3 was pipetted to the samples. The samples were incubated on ice for 10 min. After 10 min centrifugation at 12,000 x *g* the supernatant was pipetted onto a column and mixed with 200 µL of binding matrix. Following a centrifugation step (30 s, 12,000 x *g*), the flow through was discarded and the plasmid DNA that was bound to the matrix was washed twice using 500 µL of washing buffer. After the first washing step, the columns were centrifuged for 30 s and succeeding the second washing step the columns were centrifuged for 2 min at 12,000 x *g*. In the final step the plasmid DNA was eluted from the columns by adding 100 µL of H_2O, TE or EB-buffer to the columns. These were centrifuged for 1 min at 12,000 x *g* and the plasmid DNA was collected in a 1.5 mL reaction tube. The plasmid DNA was stored at -20 °C.

Resuspension solution (1):	50 mM glucose 25 mM Tris-HCl (pH 8.0) 10 mM EDTA (pH 8.0) 20 µg/mL RNase A	Alkaline lysis solution 2:	0.2 N NaOH 1% SDS
Binding matrix:	5.3 M guanidine HCl 20 mM Tris (pH 8.0) 0.15 g/mL diatomaceous earth	Washing buffer:	20 mM Tris-HCl (pH 8.0) 2 mM EDTA (pH 8.0) 0.2 M NaCl 50% EtOH
Neutralization solution 3:	5.3 M guanidine-HCl 0.7 M KAc (pH 5.0)		

2.4.3 Nucleic acid analysis and modification

2.4.3.1 Molecular cloning

In the course of this work two types of methods were applied to insert a gene or promotor sequence into an entry vector. Either via ligation of a DNA-fragment here referred to as classical cloning, or via site-specific recombination using the gateway® cloning methodology (Invitrogen, Karlsruhe).

2.4.3.1.1 Classical cloning

The major steps for classical cloning include the amplification of a DNA sequence using a proof reading polymerase (chapter 2.4.3.8.2), the linearization of an entry vector via restriction digestion (chapter 2.4.3.3), the preparation of the DNA fragment (insert) via restriction digestion and the ligation of the insert into the entry vector (chapter 2.4.3.5). Competent *E. coli* cells (TOP10, DH5α) were transformed with the ligated DNA (chapter 2.2.4.1) and positive clones were selected on LB-plates with the corresponding antibiotics. *E. coli* colonies were then tested for the presence of the specific insert via colony PCR (chapter 2.4.3.8). After cultivation of the bacteria in liquid LB-medium (chapter 2.2.2.2), the plasmid DNA was purified (chapter 2.4.2) and digested with restriction enzymes. To check for mutations the insert was sequenced and compared to the published gene sequences found among others in the public databases such as TAIR (*158*) and ARAMEMNON (*159*). In the course of this work DNA fragments were cloned into the entry vectors pENTR1A and pENTR4.

The restriction enzymes used for cloning in this work were *Eco*RI, *Not*I, *Bam*HI and *Nco*I.

2.4.3.1.2 Gateway

Entry clones were also produced via gateway® cloning (Invitrogen, Karlsruhe, D). This included the amplification of a PCR fragment with primers containing *att*B sites. The *att*B-site-PCR fragment was then recombined and inserted into an entry vector containing *att*P sites, such as pDONR207. The expression clones (see Table 1), constructed in this work, were assembled with the help of the gateway technology via LR reactions. During this process the *att*L sites of the gateway cassette of the entry clone recombine and insert into the expression vector, which contains the *att*R sites. The BP or LR reactions were incubated for 12 – 16 h and room temperature. The complete reaction mix was then transformed into competent *E. coli* cells (2.2.4.1).

BP:	LR:
10 – 100 ng *att*B-PCR product	12.5 – 37.5 ng entry clone (pENTR1A, pENTR4, pDONR207)
37.5 ng donor vector (pDONR207)	37.5 ng destination vector (pGWBs, pAMPAT)
1 µL BP 5 x clonase buffer	0.5 µL LR Clonase™ II enzyme mix
1 µL BP clonase enzyme	
ad 5 µL with TE buffer (pH 8.0)	ad 2.5 µL with TE buffer (pH 8.0)

2.4.3.2 Determination of nucleic acid concentration

Nucleic acid concentrations were determined using the NanoDrop® ND-100 Spektrophotometers (Nanodrop Technologies, Wilmington, US), via determination of the optical density (OD). The nucleic acid concentration was determined at a wavelength of λ = 260 nm. Calculating the ration between λ = 260/280 nm indicates the purity of the nucleic acid. The value of pure nucleic acid should be between 1.8 and 2.0. The measurements were performed as described by the manufacturer.

2.4.3.3 Restriction Digestion

Two forms of restriction digestion protocols were used. Plasmid DNA was digested either as a control digestion or for cloning applications.

Control DNA digestion:
0.2 – 1.5 µg plasmid DNA
0.4 µL enzyme I (10 u/µL)
(0.4 µL enzyme I (10 u/µL))
0.1 vol of 10 x restriction buffer
H_2O_{bidest} ad 20 µL

Preparative DNA digestion:
0.5-3 µg plasmid DNA/PCR fragment
1-1.5 µL enzyme 1 (10 u/µL)
(1-1.5 µL enzyme 1 (10 u/µL))
0.1 vol of 10 x restriction buffer
H_2O_{bidest} ad 30 - 40 µL

The digestion reaction-cocktail was incubated for at least 1 h at the temperature recommended by the supplier for the specific enzymes. After the restriction digestion, the DNA was separated via electrophoresis in a 1 – 1.5% agarose gel (2.4.3.9).

2.4.3.4 Dephosphorylation

Linearized vectors were dephosphorylated prior to the ligation reaction. This was especially important when the restriction sites were blunt end or the vector was only digested with one restriction enzyme. The dephosphorylation of the vector ends inhibits self-ligation of the vector. For the dephosphorylation reaction the shrimp alkaline phosphatase (SAP) was used as described by the manufacturer (MBI Fermentas GmbH, St Leon-Rot, D).

2.4.3.5 Ligation of DNA fragments

DNA fragments were ligated using the T4 DNA Ligase (MBI Fermentas GmbH, St Leon-Rot, D). For cloning purposes the PCR fragments were ligated into the vector as recommended by the manufacturer:

20 – 100 ng linearized vector
1:1 – 1:5 molar ratio of insert DNA:vector
2 µL 10 x T4 DNA ligase buffer (5 u/µL)
1 µL T4 DNA ligase
H_2O_{bidest} ad 20 µL

The ligation reaction-cocktail was incubated for 12 to 16 h at 4 °C and then directly transformed into competent *E. coli* cells (chapter 2.2.4.1).

2.4.3.6 Genomic DNA digestion using DNase-I in RNA samples

Genomic DNA contaminations in RNA samples were removed by digestion with DNase-I (DNase-free-kit, Ambion®, Applied Biosystems, Darmstadt, D) before the

RNA was further processed for analysis. The reaction-mixture consisted of the following components:

5 µg RNA 2.5 µL 10 x buffer DNase-I 0.5 µL DNase-I (2 u/µL)
$H_2O_{ultrapure}$ ad 25 µL

After 30 min incubation at 37 °C, the DNase-I was inactivated by adding 2.5 µL of DNase-I inactivation reagent. The samples were mixed thoroughly during a 2-min incubation time at room temperature. The inactivation reagent was removed from the samples by centrifugation (1.5 min, 12,000 x *g*, Minispin, Eppendorf). The supernatant was transferred into a new reaction tube. The DNase treated RNA samples were stored at -80 °C.

2.4.3.7 Reverse transcription

Reverse transcription (RT) is the synthesis of complementary DNA (cDNA) from a messenger RNA (mRNA) template. A viral RNA-dependent-DNA-polymerase, the reverse transcriptase, is necessary for this process. The RT reactions in this work were performed with the BioScript™ from Bioline, which is derived from the Moloney Murine Leukaemia Virus. The oligo dT_{18} primer is complementary to the poly(A) tail of mRNAs. Thus all mRNAs containing a poly (A) tail can putatively be transcribed into cDNA. The RT consisted of the following components:

8 µL DNase-I treated RNA (1.6 µg) (see chapter 2.4.3.6) 1 µL oligo dT_{18} primer (0.5 M) 4 µL 5 x buffer (Bioline®) 0.25 µL BioScript™ 200 u/µL (Bioline®) 1 µL dNTPs (10 mM)
Σ 5.75 µL $H_2O_{ultrapure}$

1) 25 °C → 8 min 2) 42 °C → 75 min 3) 70 °C → 15 min 4) 4 °C → ∞

The cDNA was stored at -20 °C until further use.

2.4.3.8 Polymerase Chain Reaction (PCR)

During the PCR, DNA fragments are amplified *in vitro* using a DNA-dependent-DNA–polymerase (*160, 161*). The amplification of a DNA region is among others applied as a detection and/or quantification method of a certain DNA region or fragment, for mapping of an insertion site and for cloning applications. The DNA polymerase extends short double stranded DNA fragments with 3' hydroxy-termini. Therefore, the PCR reaction can only start, when the appropriate primer oligonucleotides with free 3' hydroxy-termini are present. In a standard PCR reaction, the samples are heated to 92 – 98 °C, depending on the type of DNA polymerase. During this process, the double stranded DNA is denatured and single DNA strands are formed. In the next step, the samples are cooled down to the annealing temperature of the primers. After the primers anneal to the complementary sites on the template DNA, the extension of the primers takes place at 72 °C, which is the optimal temperature for the DNA polymerase. The extension time is dependent on the DNA fragment size that is to be amplified. This process of denaturation, annealing and extension is repeated in 25 – 40 cycles. In a final extension step at 72 °C incomplete amplicons are extended. Different types of DNA polymerases exist. In this work standard PCR reactions such as colony PCRs for the detection of positive insert containing bacterial clones, the analysis of plant genotypes, as well as semi quantitative PCRs were performed with dream taq polymerase (MBI Fermentas, St.Leon-Rot, D). During mapping of the TAMARA lines taq polymerase from Qiagen (Hilden, D) was also applied. For cloning applications, the proof reading polymerase from MBI Fermentas was used in order to obtain amplicons with a minimum of mutations.

2.4.3.8.1 Standard PCR reaction

Reaction mixture:	PCR program:
1 µL template DNA (5 pg- 0.5 µg) 0.4 µL primer 1 (10 mM) 0.4 µL primer 2 (10 mM) 0.4 µL dNTPs (10 mM) 2 µL 10 x dream taq buffer 0.04 µL dream taq (5 u/µL)	1) 96 °C → 10 min 2) 96 °C → 30 s 3) Tm- 5 °C → 30 s 4) 72 °C → 1 min/kb → 24 – 35 x to step 2 5) 72 °C → 10 min 6) 16 °C → ∞
Σ H_2O_{bidest} ad 20 µL	

2.4.3.8.2 PCR reaction with proof reading polymerase

For cloning purposes it was necessary to obtain amplicons with a minimum of mutations. Therefore, a DNA polymerase with proof reading activity (3' => 5' exonuclease activity) was applied.

Reaction mixture:	Program:
50 pg – 1 µg template DNA 5 µL 10 x *Pfu* buffer with $MgSO_4$ 1 µL dNTPs (10 mM) 2.5 µL primer 1 (10 mM) 2.5 µL primer 2 (10 mM) 1 µL *Pfu* DNA polymerase (2.5 u/µL)	1) 95 °C → 1 – 3 min 2) 95 °C → 30 s 3) Tm-5 °C → 30 s 4) 72 °C → 2 min/kb → 34 x to step 2 5) 72 °C → 5-15 min 6) 16 °C → ∞
Σ H_2O_{bidest} ad 50 µL	

2.4.3.9 Agarose gel electrophoresis

The agarose gel electrophoresis is a method for separating of nucleic acids by size. Since the nucleic acid is negatively charged, it migrates through the pores of the agarose gel towards the anode, once an electric field is applied. Typically a 1 – 1.5% agarose gel was subjected to a voltage of 70 – 120 V and a current of ca. 150 mA. The nucleic acid was stained with ethidium bromide (EtBr), which intercalates into nucleic acids. Upon exposure to UV-light, the EtBr fluorescents, the fluorescence is intensified when EtBr is bound to nucleic acid. In this way a nucleic acid band in the gel can be visualized. The agarose was suspended in 1 x TAE buffer and heated until completely dissolved. After the agarose solution had cooled down to a temperature of 50 – 60 °C, 0.5 µg/mL of EtBr was added. The agarose was poured into a plastic form and the appropriate comb was used to form pockets for samples in the gel. After the agarose gel was solid, it was transferred into a gel chamber filled with 1 x TAE buffer. The nucleic acid samples, complemented with loading buffer, where pipetted into the pockets of the gel. For the determination of the size of the DNA fragments 5 µL of a ladder of known DNA fragments (Fig. 12) was also separated during electrophoresis. The bands were documented with a digital camera (Kaiser RT1, INTAS).

50 x Tris acetate EDTA-buffer (TAE):	2 M Tris/HAc pH 7.5 50 mM EDTA
10 x loading buffer:	20% (w/v) ficoll 400 1% (w/v) SDS 0.05% (w/v) xylene blue 0.05% (w/v) bromphenol blue 40 mM Tris / HAc (pH 7.5) 100 mM EDTA
ethidiumbromide (EtBr)	1% EtBr in H_2O

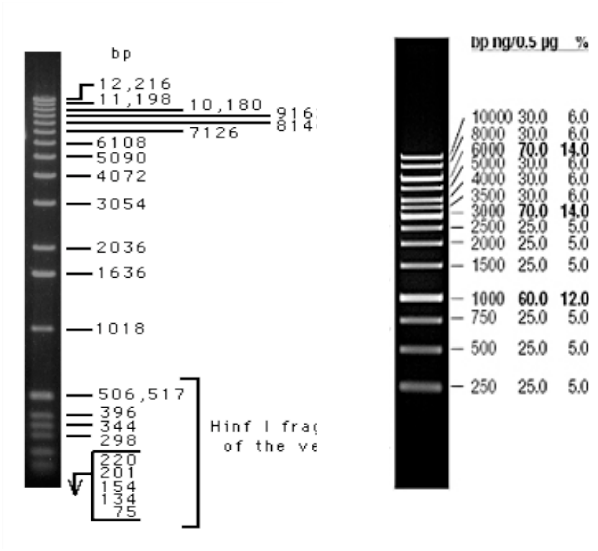


Fig. 12: DNA standards applied for gel electrophoresis
Left: 1 kb DNA ladder (Invitrogen, Karlsruhe, D), right: gene ruler 1 kb DNA ladder (MBI Fermentas St. Leon-Rot, D).

2.4.3.10 DNA purification from agarose gels

The purification of DNA fragments from agarose gels was performed using the kits E.Z.N.A® gel purification (Omega Bio-Tek, Inc, Norcross, GA, USA), QIAquick gel extraction and MinElute Gel extraction (Qiagen GmbH, Hilden, D) as recommended by the manufacturer.

2.4.3.11 Ethanol precipitation

For purifying DNA, 1/10 volume of 3 M sodium acetate (pH 5.2) was mixed with the DNA. 2.5 volumes of 100% EtOH (– 20 °C) was added to the samples and the mixture was centrifuged for 20 min at 18,000 x *g* (4 °C) (Eppendorf centrifuge 5417). The DNA pellet was resuspended in 250 µL 70% EtOH and centrifuged again for 10

min at 18,000 x *g* (4 °C). The supernatant was discarded and the pelleted DNA was air dried before it was resuspended in the necessary buffer-solution.

2.4.4 Mapping of the insertion site

The selected candidate mutants from the TAMARA population were mapped using the "En flanking"- method adapted from Varotto *et al*. (2000) (*162*). This included nested PCR reactions and the sequencing of the flanking region of the T-DNA insertion site. A high yield of pure genomic DNA was required and therefore the DNA was extracted from inflorescences using the protocol described in chapter 2.4.1.2. 50 – 100 ng of genomic DNA was digested with 2 units of the restriction enzymes (0.2 µL) *Csp*6I and *Hind*6I, in separate reactions. Alternatively the genomic DNA was digested with *Hind*III/*Eco*RI and with *Vsp*I, at 37 °C for 3 – 4 h. Inactivation of the enzymes followed for 10 min at 65 °C. 10 µL of the digested genomic DNA was ligated to the adaptors APL1732 (for *Hin*6I) and APL 1632 (for *Csp*6I) (see 2.4.4.1). 4 µL of the ligation reaction functioned as template for PCR 1, using the primer EN205R, which anneals to the adaptor sequence. For PCR 1, the taq polymerase from Qiagen was utilized. In a second PCR the PCR 1 product was used as template. In PCR 2, the primers En82as and LR26 were applied, thus the 5' end of the T-DNA and the flanking genomic DNA region were amplified. For PCR 2, the taq polymerase from Fermentas was used. The DNA bands were separated by electrophoresis and eluted from the gel using QIAquick or MinElute gel extraction kits (Qiagen, Hilden, D). The DNA fragments were sequenced using the primer LR26 or EN82as. If the digestion of the genomic DNA for mapping was performed with the restriction enzymes *Hind*III/*Eco*RI and *Vsp*I, then the adaptors Ase, Hind and Eco were attached. For PCR 2 the adaptor primer AP1 was combined with EN82as. En82as was then also used for sequencing. The mapping procedure is summarized in Fig. 13.

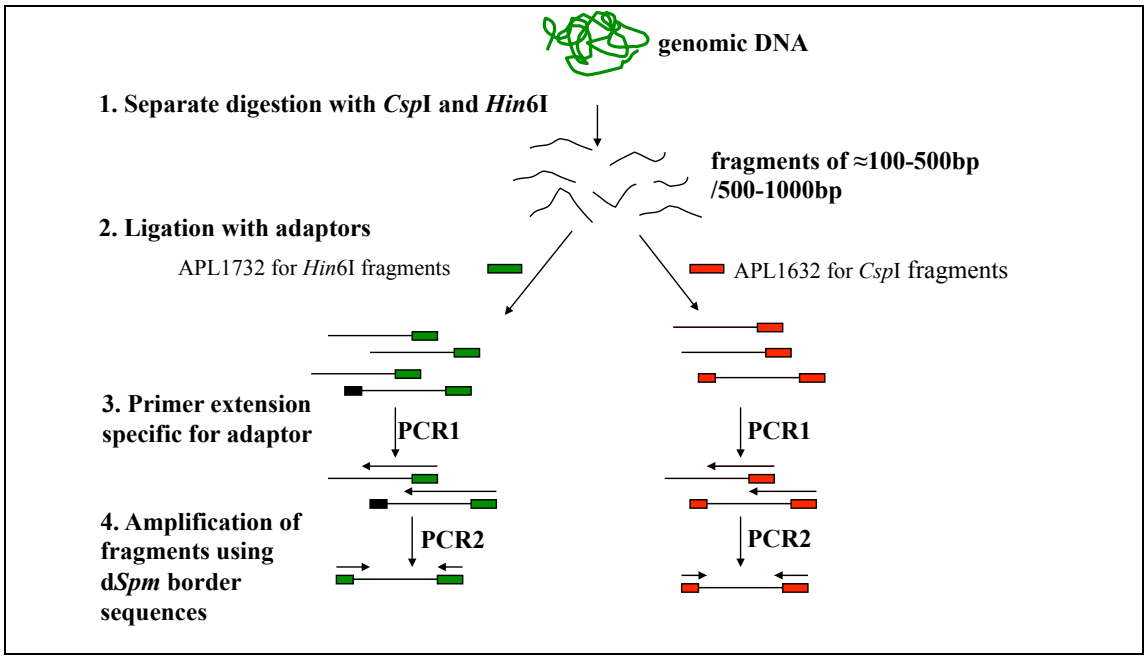


Fig. 13: Schematic summary of the mapping process.

2.4.4.1 Adaptor production

The adaptors were produced via annealing the primers LR32 and APL16 (APL 1632) or LR32 and APL17 (APL 1732). The primers were diluted 1:10 in annealing buffer (10 mM Tris (pH 7.5); 10 mM $MgCl_2$). After a 5-min incubation at 80 °C, the primers were cooled down slowly. The adaptors could be stored at -20 °C until further use.

2.4.5 Sequencing of double stranded DNA

Sequencing of double stranded DNA, after cloning or for mapping purposes, was performed by the Cologne Center of Genomics (CCG). The BigDye® terminator v3.1 was applied for all sequencing reactions, as recommended by CCG.

2.5 Human and murine hepatoma cell culture

2.5.1 Cultivation of human and murine cell lines

2.5.1.1 HepG2-GFP cells

For long-term storage, the cells were frozen in liquid nitrogen. To start the cultivation of the cells, a cell culture petri dish (cell culture dish 100 mm x 20 mm treated polystyrene, Corning Inc., NY, USA) was filled with 10 mL of RPMI (37 °C). The

frozen HepG2-GFP cells were rapidly thawed at 37 °C and immediately diluted into the RPMI medium. The cells were evenly distributed in the petri dish. For the cultivation of the cells, the petri dish containing the cells was incubated at 37 °C, 5% CO_2 and 95% air humidity (InCu safe MCO-20 AIC, SANYO, München, D).

Passaging of the cells into new medium was required every 48 – 72 h. For cell survival it is necessary for them to attach to the surface of the plastic dish. The cells can be passaged up to 20 times, before it is necessary to start a new culture from the frozen stock. For passaging the cells, the medium was removed using a vacuum pump (Vaccubrand GmbH, Wertheim, D) with a sterile glass tip. The cells were washed twice with 10 mL 1x DPBS (37 °C). In the next step the cells were removed from the plastic surface by treating them for 20 - 25 min with 1 mL of the enzyme accutase (PAA). Alternatively the cells were treated for a minimum of 5 min but maximum 15 min, with 1 mL of trypsin-EDTA at 37 °C. 3 mL of RPMI was added to the detached cells before they were separated by carefully pipetting them up and down. The 4 mL cell suspension was divided into ½, ¼ and $^1/_8$ volumes and transferred into new petri dishes, each containing 10 mL of RPMI. The cells were distributed evenly and incubated at 37 °C, 5% CO_2 and 95% air humidity.

RPMI:	500 mL of RPMI 1640 (Invitrogen) with: 15% FBS (90 mL) 2 mM L-glutamine (Invitrogen) (6 mL of 200 mM) 100 U penicillin/100 µg streptomycin (Invitrogen) (6 mL)
Dulbecco's Phosphate-Buffered Saline: (DPBS) (Gibco®,invitrogen)	1.37 M NaCl 0.027 M KCl 0.08 M Na_2HPO_4 0.015 M KH_2PO_4
Trypsin-EDTA (Gibco®, invitrogen)	
Accutase (PAA)	

2.5.1.2 Hepa1c1c7-Lux cells

The Hepa1c1c7-Lux cells were passaged as described for the HepG2-GFP cells in chapter 2.5.1.1, with the following alterations. The cells were cultivated in α-MEM medium. The 4 mL cell suspension was divided into $^1/_4$, $^1/_8$ and $^1/_{16}$ volumes.

α-MEM:	500 mL of α-MEM (Invitrogen) with: 15% FBS/FCS (90 mL) 2 mM L-glutamine (Invitrogen) (6 mL of 200 mM) 100 u/mL penicillin/ 100 µg/mL streptomycin (Invitrogen) (6 mL) 417 µg/mL Geneticin G418 (Invitrogen) (5mL)

2.5.2 Production of DMSO stocks of cell culture for long-term storage of HepG2-GFP and Hepa1c1c7-Lux cells

For DMSO stocks, the cells were washed twice in 10 mL 1 x DPBS. The cells were detached from the surface via a 5 min treatment at 37 °C with 1 mL of trypsin-EDTA. 2 mL of medium was added to the trypsinated cells. The following steps were performed on ice. After carefully mixing the medium with the cells, they were transferred into a 15 mL sterile plastic tube. The plate was washed oncemore with 5 mL of medium, which was also transferred to the plastic tube. The cells were centrifuged at 130 x *g* (Eppendorf centrifuge 5810R) and the pellet was resuspended in 3 mL freeze medium. The aliquots of 1 mL each were pipetted into 2 mL-cryo tubes. The cells were frozen for 1 – 2 weeks at -70 °C. In order to freeze the cells slowly, they were wrapped in paper and stored in a thick Styrofoam box. Finally the cells were transferred from -70 °C to liquid nitrogen for long-term storage. During the freezing process to -70 °C, it was necessary that the cells were not moved to avoid liquid crystallization within the cells and their disruption.

Recovery freezing medium (Gibco®, Invitrogen)

2.5.3 Screening assays

2.5.3.1 Preparation of plant methanol extracts

30 mg of leaf material was harvested and immediately frozen in liquid nitrogen. The plant material was homogenized using the tissue lyser (Qiagen, Hilden, D). 300 µL 80% methanol was added and the samples were homogenized again. The samples were centrifuged at 18,000 x *g* for 20 min (4 °C) (Eppendorf centrifuge 5417). The supernatant was transferred into a new reaction tube. The methanol extracts were vacuum dried (Eppendorf concentrator 5301) and the completely dried pellets were resuspended in 100 µL DMSO.

2.5.3.2 GFP fluorescence assay with HepG2-GFP cells

HepG2-GFP cells were detached from the surface of the cultivation plate as described in chapter 2.5.1.1. The amount of cells per microliter was determined using a counting chamber (Haemacytometer, NEUBAUER GmbH, Lauda-Königshofen, D). 50,000 cells were distributed into each well of a black 96-well plate (Costar, 3916, Corning inc. NY, USA) and the necessary amount of RPMI-medium was added to reach a total volume of 200 µL per well. The cells were incubated for 24 h at 37 °C, 5% CO_2 and 95% air humidity (InCu safe MCO-20 AIC, SANYO, München, D). 2 µL of the prepared DMSO plant extracts, diluted 1:5 in RPMI-medium, was added to each well. The cells were again incubated for 24 h before the GFP-fluorescence was quantified. For the GFP-quantification the medium was carefully removed from each well using a vacuum pump. 200 µL of DPBS with EtBr (100 µg/mL) was added to each well. After 20 min incubation in the cell culture incubator, the buffer was removed using the vacuum pump (Vaccubrand GmbH, Wertheim, D). After a washing step with 200 µL of DPBS, again 200 µL of DPBS were added to each well and the GFP fluorescence was determined using the microplate reader (Infinite 200, Tecan GmbH, Crailsheim, D). The cells were stained with EtBr in order to quantify the amount of cells per well. The GFP fluorescence was determined at λ = 485/530 nm and the EtBR at λ = 485/612 nm. For analysis the data was exported from the software Magellan into excel. The GFP fluorescence was divided by the EtBr fluorescence per well.

2.5.3.3 Luciferase assay with Hepa1c1c7-Lux cells

2.5.3.3.1 Preparation of the 96 well plates

The Hepa1c1c7-Lux cells were detached from the cultivation plate as described in chapter2.5.1.2. Using the counting chamber to determine the amount of cell per microliter, 20,000 cells were distributed into each well of clear 96 well microtiter plate (Costar, 3596, Corning inc. NY, USA) and α-MEM medium was added to reach a final volume 200 µL per well. The plates with the prepared cells were incubated for 24 h. 2 µL of 1:5 in medium or DPBS diluted DMSO plant extracts were pipetted into each well to induce the cells. After 24 h of incubation the luminescence and protein content was determined for each well, as described in the following chapters.

2.5.3.3.2 Luciferase assay

The medium was removed from each well using a vacuum pump (Vaccubrand GmbH, Wertheim, D). 200 µL of DPBS was added to each well for washing the cells. After removing the buffer with the vacuum pump, 40 µL of 1 x passive lysis buffer (Promega, Mannheim, D) was added to each well. The cells were incubated on a shaker for 15 min at 600 rpm (MHR23, HLC, Ditabis, Pforzheim, D). The cells were pelleted for 15 min at 2900 x *g* (Eppendorf centrifuge 5810R). 15 µL 2 mM EDTA was pipetted into each well of a white 96 well micro titer plate. 10 µL of the supernatant of the lysed cells was added into the wells. The luciferase buffer was prepared freshly for each screening and stored on ice. The luciferase buffer was automatically pipetted using the microplate reader (Infinite 200, Tecan GmbH, Crailsheim, D). 100 µL of buffer was added to each well, the luminescence was determined. The program I- control settings for the microplate reader were as following:

amount: 100 µL per well; speed: 200 µL/s; wait: 2 s; luminescence; integration: 10000 ms; attenuation: none

luciferase-assay buffer:	1 x Luciferase-assay buffer 15 mM DTT 1 mM ATP 0.5 mM Luciferin	2 x luciferase assay buffer:	100 mM Tris-HCl (pH 7.8) 20 mM $MgCl_2$ 2 mM EDTA

2.5.3.3.3 BCA-assay

The BCA™ protein assay kit (Pierce, Rockford, IL, USA) was used for determining the protein content of all lysates. The protein content of each well was used to relate assay results to cell densities. 15 µL of 2 mM EDTA was pipetted into a clear 96 well micro titer plate. 10 µL of the supernatant of the lysed cells was added. The solutions A and B of the BCA kit were mixed in a relation of 50:1 and 200 µL of solution AB was pipetted to each sample. The samples were incubated for 30 min at 37 °C, a purple coloring could be observed depending on the protein content. After the plates cooled down to room temperature, the samples were measured at λ = 562 nm in the microplate reader (Infinite 200, Tecan GmbH, Crailsheim, D) using the program I-control. The protein content per well was calculated using a standard curve that was obtained by testing different bovine serum albumin (BSA) concentrations.

2.5.3.3.4 Quinone reductase assay

With the quinone reductase (QR) assay (*117*) the activity of an endogenous phase II detoxification enzyme in the Hepa1c1c7-Lux cells was quantified. The basis of the assay is the reduction of menadione to menadiol by NADPH/NADH, which is catalyzed by the quinone reductase. 3-(4,5-dimethylthiazo2-yl)-2,5-diphenyltetrazolium bromide (MTT) is then nonenzymetically reduced by menadiol. The reduction of MTT results in a blue color formation, which can be measured at λ = 595 nm. The supernatant, 10 µL, of the lyzed cells from the luciferase assay (chapter 2.5.3.3.2) were pipetted into a 96-well microtiter plate and mixed with 200 µL of QR assay solution. The samples were incubated for 25 min at room temperature before the absorption was determined using the microplate reader (Infinite 200, Tecan GmbH, Crailsheim, D)

QR-solution:	25 mM Tris-HCl (pH 7.4) 30 µM NAD 5 µM FAD 50 µM menadione 1 mM glucose-6-phosphate 0.07% (w/v) BSA 0.03% (w/v) MTT 0.01% (v/v) Tween-20 1 u/mL glucose-6-phosphate dehydrogenase (*Leuconostoc mesenteroides*)

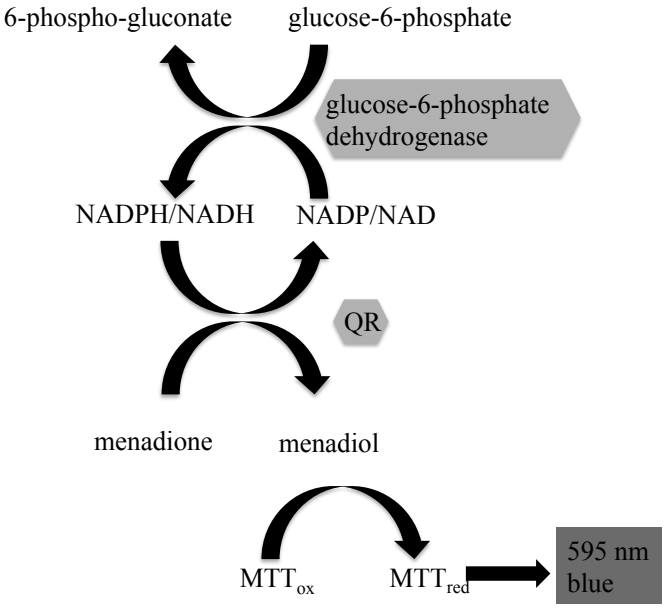


Fig. 14: Schematic overview of the reactions during the quinone reductase assay. Adapted from (*117*).

2.5.3.3.5 Analysis of the assay results

The luminescence per well or the absorption determined during the QR assay was divided by the protein content of the respective well. The protein content in µg/mL was determined via a standard curve using BSA. Typically, three technical replicates were made for each mutant. The mean luminescence per protein content ($\bar{x}$) of the technical replicates (n) was calculated for each sample and the standard error ($SE_{\bar{x}}$) was determined using the following formula:

$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$, with s = standard deviation.

During the screening, each 96-well plate included 3 technical replicates of a Col-*0* sample used as a standard for determining the EpRE inducing potential of a mutant extract. To quantify the induction potential, the mean mutant luminescence per protein (µg/mL) was divided by the mean Col-*0* luminescence per protein (µg/mL). A mutant leading to a ratio above 1 was interpreted as a potential chemoprotective mutant. In some cases, the determined Col-*0* value was unusually high in comparison to all other screened samples of the plate. The Col-*0* plant most likely was subjected to unintented stress and therefore the value was seen as an outlier. Instead the mutants were compared to average Col-*0* values of several other screenings in order consider mutants of the screen as putative chemoprotectants.

2.6 Plant biotechnological methods

2.6.1 Floral transformation

Stable transgenic *A. thaliana* plants were generated via vacuum infiltration of the inflorescences with *A. tumefaciens* containing the gene of interest. The gene of interest was integrated into the Arabidopsis genome via T-DNA insertion. *A. tumefaciens* of the strain GV3101 pMP90 or GV3101 pMP90RK were cultivated on a shaker in 5 mL YEB-medium containing selective antibiotics. After ca. 16 h of cultivation at 28 °C and 220 rpm shaking intensity the bacteria were transferred into 400 mL fresh YEB medium with antibiotics. The bacteria were grown until they reached an OD_{600} = 1.0. After centrifugation of the bacteria for 15 min at 3,500 x *g* (Eppendorf centrifuge 5810R), they were resuspended in 1/3 vol. infiltration medium

(ca. 130 mL). After transferring the bacteria into a glass dish, the inflorescences of Col-*0* plants were dipped into the suspension. Before the transformation process, all siliques were removed from the flowering plants. A vacuum was applied using a water-jet vacuum pump. After 10 -15 min of vacuum infiltration, more bacteria were applied onto the inflorescences using a pipette. The plants were covered with a plastic lid and 12 – 16 h later more *A. tumefaciens* were pipetted onto the inflorescences. The plants were kept covered with a plastic lid for 2 days and transferred into the green house.

Infiltration medium:	2.56 mM MES/KOH (pH 5.8) 2.2 g/l Murashige & Skoog salts (Duchefa M0221) 146 mM Saccharose 10 µL 6-benzymaminopurin (1 mg/mL in 1 N NaOH) 0.1% Tween 20

2.6.2 *Arabidopsis* root cell culture

The *Arabidopsis* root cells were cultured in liquid AT medium at 22 °C and 120 rpm (shaker: Multitron, Infors HT, Einsbach, D). The cells were diluted 1:3 - 1:5 every 7 days depending on the growth rate of the cells.

Arabidopsis thaliana medium (AT):	4.4 g/L Murashige & Skoog salts (Duchefa M0221) 3% sucrose 4 mL/L B5 vitamin mixture (1 g/L nicotinic acid, 1 g/L pyridoxin-HCl, 10 g/L thiamin HCl, 100 g/L myo-inositol) 0.5 mL 2,4 D (2 mg/mL) (pH 5.8)

2.6.3 Transient protein expression in *Arabidopsis* root cell culture modified from (*163*)

Supervirulent (SV) agrobacteria containing the construct of interest were grown for 24 h at 28 °C and 220 rpm in 5 mL liquid YEB medium in 15 mL reaction tubes (shaker: Multitron, Infors HT, Einsbach, D). In addition, the antisilencing strain P19 was cultivated in a separate reaction tube. The Arabidopsis Col-*0* root suspension culture was diluted 1:4 or 1:5, depending on the density of the pre-culture, in fresh *A. thaliana* (AT)-medium (room temperature) prior to transfection. The SV and P19

agrobacteria were centrifuged at 4,000 x *g* for 7 min (Eppendorf centrifuge 5810R). The bacterial pellet was resuspended in 500 - 1000 µL of AT-medium. 3 mL of diluted Arabidopsis cell culture were pipetted into each well of a sterile 6 well plate (Corning Inc. NY, USA). 25 µL of SV and P19 agrobacteria were added to each well. The plates were wrapped in aluminum foil to keep the cells dark and to prevent the evaporation of AT-medium. The plant cell culture was cultivated at 22 °C and 120 rpm for 4 - 6 days before the GFP or ß-glucuronidases (GUS) activity was analyzed.

2.6.4 Transient expression of protein in *N. benthamiana* and *A. thaliana* leaves

Two months old *N. benthamiana* plants or 6 - 8 weeks old *A. thaliana* plants, grown under short day conditions, were used for transient expression of proteins. The SV agrobacteria were grown for 2 days on YEB-plates with the respective antibiotics. The bacteria were the cultivated for 24 h at 28 °C (220 rpm) in 50 mL liquid YEB-medium. The bacteria were centrifuged for 7 min at 4000 x *g* and resuspended in 1x agromixsolution. The OD_{600} was adjusted to 0.7- 0.8 with the agromixsolution. The agrobacteria with the gene of interest were mixed in equal amounts with the antisilencing strain P19 (*154*). The agrobacteria were incubated for 2 to 6 h at room temperature and then infiltrated into the abaxial side of the leaves. Prior to infiltration, the plants were watered and kept in the dark for 2 to 6 h. The infiltrated plants were covered with a lid and 24 h later transferred into the green house. If the expressed protein was fused to a fluorescent marker for subcellular localization studies, the plants were analyzed 5 to 6 days after agroinfiltration.

Agromixsolution:	10 mM $MgCl_2$+H_20 100 mM MES (pH 5.6) 0.15 mM acetosyringone

2.7 Analytical methods

2.7.1 Real time PCR

After RNA preparation and reverse transcription, the expression of a gene in a specific plant tissue can by quantified by real time PCR. The real time PCR was performed using the fluorescence dye SYBR green, which binds mainly to double stranded DNA, absorbing blue light at a wavelength of $\lambda = 497$ nm and emitting green

light at a wavelength of λ = 520 nm. The emission is used to quantify the amount of double stranded DNA in a sample. After each PCR cycle the fluorescence is measured. In this way the cycle threshold (Ct) value can be determined, which is the cycle during which the fluorescence is significantly stronger than the background signal (exponential phase of the PCR). The cDNA was then diluted 1:4 – 1:8 with $H_2O_{ultrapure}$. During the real time PCR the amount of cDNA in each well was quantified via amplification of the housekeeping gene actin2, which was used as an internal reference gene. The gene specific real time PCR primers were designed using the software primer express (Applied biosystems), with a T_m of 60 °C and an amplicon length of 70 – 150 bp. Duplicate reactions were set up for each cDNA sample. The real time PCR reaction was performed using 7300 System Software Instrument (Applied Biosystems, Darmstadt, D). The SYBR green PCR mix contains a heat activated DNA polymerase, ultra pure dNTPs, 6 mM $MgCl_2$, the internal reference ROX and SYBR green dye (Applied Biosystems, Darmstadt, D).

PCR reaction mix:	Program:
2 µL diluted cDNA template 1.25 µL primer forward (10 mM) 1.25 µL primer reverse (10 mM) 10 µL 2 x SYBR green PCR mix 5.5 µL $H_2O_{ultapure}$	1) 50 °C → 2 min 2) 95 °C → 10 min 3) 95 °C → 15 min 4) 60 °C → 1 min 5) 95 °C → 15 s
Σ 20 µL	6) 60 °C → 1 min → 39 x to step 2 7) 95 °C → 15 s 8) 60 °C → 15 s

A dissociation stage was added at the end of the PCR program to check for primer dimers and contaminations in the minus template control. The amount of DNA was calculated via the relative quantification method:

dCt-value = Ct(target gene) – Ct(internal reference gene)

ddCt-value = dCt – dCt (control sample)

relative quantification (RQ) = 2^{-ddCt}

Alternatively to the SYBR® green ready mix, Brilliant SYBR® Green QPCR Core Reagent kit (Agilent Technologies, Inc) was applied. In this case dream taq polymerase (MBI Fermentas, St-Leon-Rot, D) was used as DNA polymerase.

PCR reaction mix:
2 µL diluted cDNA
2 µL 10 x buffer for dream Taq
1 µL SYBR (1:3000)
3.2 µL glycerol (50%)
0.6 µL DMSO (100%)
0.2 µL dream Taq
0.4 µL ROX
0.4 µL Primer 1 (10 mM)
0.4 µL Primer 2 (10 mM)
0.4 µL dNTPs (10 mM)
9.4 µL $H_2O_{ultrapure}$
Σ 20 µL

2.7.2 Microscopy

The GFP fluorescence was observed using a fluorescence microscope (Leixa DMRB, Leica DMRA2) or a confocal microscope (LSM700, Zeiss). Leaf material was cut into small sections of ca. 5 -10 mm x 5 -10 mm using a scalpel. The sections were placed on a microscope slide with the abaxial side of the leaf facing up. Water was dropped onto the section and the sample was covered with a cover glass making sure to exclude any air bubbles. Alternatively 100 – 300 µL of *A. thaliana* cultured cells were pipetted onto the microscope slide. The GFP fluorescence is excited at λ = 488 nm and detected at λ = 510 – 520 nm.

Table 8: Filters used for microscopy

Filter	Excitation filter	Emission filter	Manufacturer
DAPI	BP340-380	LP435	Leica 512804
EGFP_LP	BP470/40	LP500	AHF
ET-EGFP (no chlorophyll)	ET470/40	ET525/50	AHF 46-002

2.7.2.1 Brefeldin A treatment of *A. thaliana* cultured cells

To confirm a subcellular localization of a protein fused to GFP to the endomembrane system or Golgi, the *A. thaliana* cultured cells were treated for 2 h with 100 µg/mL Brefeldin A (BFA) at 22 °C. The treated cells were then observed using the fluorescence microscope.

2.7.3 Metabolic profiling

The profiling of plant secondary metabolites was carried out by our collaborating partners from IPB-Halle (Stephan Schmidt, Christoph Böttcher, Dierk Scheel) using the method published by von Roepenack-Lahaye (2004) (*164*), De Vos *et al.* (2007) (*124*), Böttcher *et al.* (2008) (*129*).

2.7.4 Extraction of glucosinolates

70 - 100 mg of plant material was harvested and frozen in 2 mL tubes, in liquid nitrogen. The plant material was lyophilized for 16 – 20 h. For the equilibration of the ion-exchange-columns, 0.1 g of Sephadex A per sample (DEAE A25120, Sigma-Aldrich, Seelze, D) was incubated in 2 mL of 0.5 M acetic acid buffer (pH 5.0) for 12 – 16 h at room temperature. After lyophilization the dry weight of the plant material was determined and the dried leaves were homogenized using the tissuelyzer (1min, 25 rpm). 1 mL of 80% MeOH was added to each sample. Finally 20 µL of 4 mM benzyl glucosinolate standard were added. The samples were homogenized for another minute and centrifuged for 10 min at 20,000 x *g*. The supernatant was transferred into a fresh 2 mL tube. To the pellet 1 mL 80% MeOH was added and the plant material was homogenized a second time. After centrifugation (10 min, 20,000 x *g*), the supernatant was added to the 2 mL tube. The columns were prepared by placing a small amount of cotton into each glas pipette, then 2 mL of the Sephadex A solution were added. The collected supernatant was transferred into the columns. The columns were washed 5x with 2 mL of HPLC grade water and twice with 0.02 M acetic acid NaOH buffer (pH 5.0). 50 µL of sulfatase (*Helix pomatia*, Sigma, EC 3.1.6.1) were mixed with 450 µL of 0.02 M acetic acid NaOH buffer (pH 5.0) for each sample. Each column was loaded with 500 µL of the prepared sulfatase solution. After the sulfatase had entered the column matrix, they were sealed with white pipette tips (sealed tips) and parafilm. In addition the top of the columns was sealed with parafilm. The samples were incubated for 12 – 16 h at room temperature. The glucosinolates were then eluted from the columns with 6 mL of HPLC grade water. The eluted glucosinolates were dried using the vacuum dryer. This step took about 24 h. The dried pellets were solved in 300 mL HPLC grade water. This solvent was centrifuged for 5 min at 20,000 x *g*, before the supernatant was transferred into the suitable glas tubes, necessary for measurements using the Acquity UPLC-System

(Waters, Eschborn). During the measurements, the metabolites were separated with a linear gradient of acetonitrile and a flow rate of 0.225 mL/min.

time (min)	0	1	5.5	9.5	9.6	9.7	9.9	10
A =10% acetonitrile	100%	100%	75%	53%	5%	5%	100%	100%
B =90% acetonitrile	0%	0%	25%	47%	95%	95%	0%	0%

The desulfoglucosinolates eluted from the column BEH-C18 (1,7 µm; 2,1 Å~ 150 mm; Waters, Eschborn, D) and detected with UV light, at a wavelength of λ = 229 nm. The amount of glucosinolates was calculated per g of leaf dry weight (see Formula 1)

Formula 1: Calculation of glucosinolate (GLS) content

$$\text{GLS in µmol/g dry weight} = \frac{\text{area GLS} * \text{response factor GLS} * \text{IS (µMol)}}{\text{area IS} * \text{response factor IS} * \text{dry weight (g)}}$$

2.7.5 GUS-Assay

As preparation for the GUS assay, the premix solution II and the X-Glu-solution were mixed in a ratio of 19:1 (for 10 mL working solution: 9.5 mL premix II + 0.5 mL of X-glu). One to 2 mL of the prepared GUS solution were used for each leaf sample. The solution was vacuum infiltrated into the tissue. The samples were incubated over night at 37 °C. To remove the chlorophyll from the tissue, the samples were incubated in 80% EtOH at 60 °C. Samples were stored at 4 °C.

1 M $NaPO_4$	working solution:	0.1 M $NaPO_4$
0.25 M EDTA		10 mM EDTA
5 mM K-Ferricyanid		0.5 mM K-Ferricyanid
5 mM K-Ferrocyanid		0.5 mM K-Ferrocyanid
20 mM X-Glu (100 mg in 10 mL dimethylformamide)		1 mM X-Glu
10% Triton X-100 (v/v)		0.1% Triton X-100 (v/v)

3 Results

The first part of this section describes the screening process used in this work. The results of the mapping, as well as the expression analyses are presented. In the second section selected mutants were analyzed in relation to their function and the production of chemoprotective compounds.

3.1 The screening of mutants of the TAMARA population

The extracts of approximately 5,000 plants of the TAMARA population were tested for chemoprotective activity in previous screening assays using the HepG2-GFP reporter system (2.5.3.2) (*112*). After two rounds of screening, approximately 470 mutants showing a high inducing effect were selected for further analysis using Hepa1c1c7-Lux cells (2.5.3.3). The screening strategy is summarized in Fig. 15.

In the following chapters the results of the screening using the Hepa1c1c7-Lux cells are described. Furthermore, the genes detected after the mapping are categorized, in order to gain more knowledge about the functions and characteristics of genes involved in the production of chemoprotective compounds. In the last chapter mutants exhibiting phenotype under green house conditions are described.

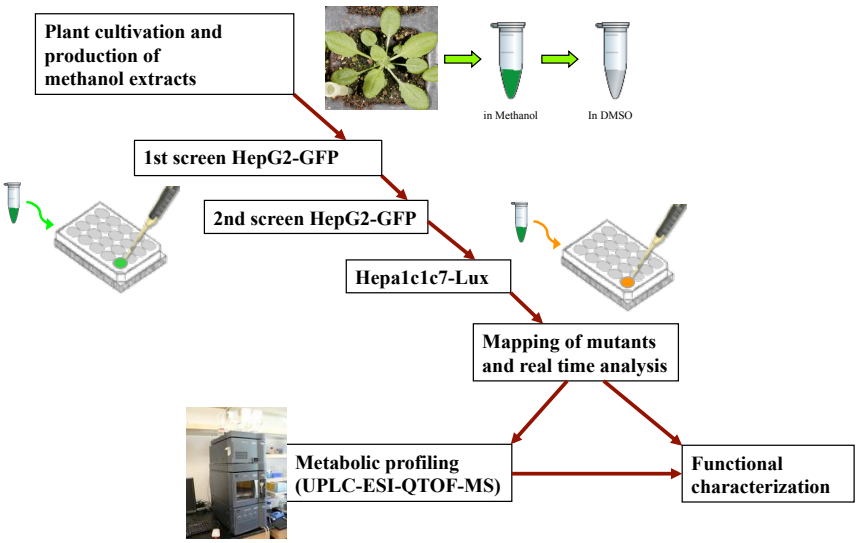


Fig. 15: Summary of the screening procedure. The plants were cultivated and methanol extracts were prepared from leaf material. The metabolites were concentrated, resuspended in DMSO and applied to HepG2-GFP cells. After the selection of candidates, these were confirmed using Hepa1c1c7-Lux cells. The confirmed mutants were mapped and the relative expression levels of genes flanking the insertion site were determined. Selected genes were further analyzed in parallel to metabolic profiling by UPLC-ESI-QTOF-MS.

3.1.1 Application of the luciferase assay system using plant extracts

Boerboom *et al*. (2006) described the development and application of the stable Hepa1c1c7-Lux reporter cell lines (EpRE(mGST-Ya)-LUX). These cells contain an EpRE-Lux construct, which enables the screening for compounds with EpRE inducing activity, by simply measuring the luminescence of the induced cells (*113*). The reporter system was applied for the analysis of different flavonoids with planar or non-planar structure.

Here, the Hepa1c1c7-Lux reporter cell lines were used to test mutant plant extracts for chemoprotective activity. Since the cells were so far only applied to test isolated compounds, the efficiency of the reporter system was assessed with different amounts of Col-*0* plant extracts. The extraction was performed with 80% methanol because in previous work, by a former bachelor student involved in this project, it was shown that this method was more efficient in comparison to aequous extraction and led to a stronger induction of HepG2-GFP cells (*165*). The induction potential of 1, 2 and 4 µL of Col-*0* wild-type plant extracts redissolved in DMSO and the DMSO dilutions of these in 1:2, 1:4, 1:8 and 1:12 were analyzed. In this experiment, concentrated extracts derived from 240 mg of leaf or inflorescence material were redissolved in 500 µL DMSO each. This corresponds to 480 µg of the original biomass of plant material per 1 µL DMSO. The applied amounts of plant extract (in µg) are summarized in Table 9.

Table 9: Summary of the applied amounts of Col-*0* plant extract.

Applied amount of extract	**Dilution**				
	1:1	1:2	1:4	1:8	1:12
1 µL	480 µg	240 µg	120 µg	60 µg	40 µg
2 µL	960 µg	480 µg	240 µg	120 µg	80 µg
4 µL	1920 µg	960 µg	480 µg	240 µg	160 µg

In Fig. 16 the results of the induction with leaf extracts are shown. At 1 µL application volume (Fig. 16A) the differences of the luminescence signal between 40, 60 and 120 µg were only slight, with intensities comparable to the untreated cells, while with 240 and 480 µg a noticeable induction of the luminescence signal could be

observed. This demonstrates an almost linear relationship between the amount of extract applied and the luminescence signal. When 2 µL of extract were applied to the system (Fig. 16B), the luminescence was induced upon treatment with 240, 480 and 960 µg of extract in comparison to those cells treated with DMSO, 80 and 120 µg of extract. The average luminescence signal was almost equal in cells treated with 480 and 960 µg of extract. When 4 µL of leaf extract were administered to the reporter cells (Fig. 16C), the application of 1920 µg extract resulted in luminescence intensity comparable to untreated cells. 960 µg extract led to a strong increase of the signal, in comparison to 1920 and 480 µg extract. In addition it could be shown that elevated amounts of DMSO had a negative effect on the induction of the cells. The results of the same process performed with extracts from Col-*0* inflorescences showed a similar pattern.

A previously described activation tagged mutant, HIGH INDOLIC GLUCOSINOLATE1-1DOMINANT (*HIG1-1D*), was used to induce the Hepa1c1c7-Lux cells. *HIG-1D* accumulates a four to six fold higher amount of I3M (*151*). One potent inducer of phase II enzymes is I3C, which is derived from I3M (*166*). In the Hepa1c1c7-Lux cell assay system a three fold elevated luminescence signal in comparison to the signal induced by wild-type plant extracts was observed. In contrast, the *hig1-1* loss of function mutant did not show an altered inducing effect when compared to Col-*0* (Fig. 18).

The data demonstrates that the EpRE based reporter system in Hepa1c1c7-Lux cells can be applied to screen plant extracts for chemoprotective activity based on the induction of phase II and detoxification enzymes that are regulated by EpRE promotor elements. In general, increased amounts of plant extract led to a higher luminescence signal. Nevertheless, a very high concentration of extract led to an inhibition of luminescence, this was most likely due to a toxic effect of compounds found in the extract. Since the aim is to screen for mutants showing higher activity than Col-*0* extracts, the applied amount of extract during the screening should enable a higher luminescence signal than induced by Col-*0*. If the applied concentration is too high, then extracts from mutants containing more phytochemicals may be toxic and therefore decrease the signal. Due to this reason the screenings were performed with 120 µg of plant extract.

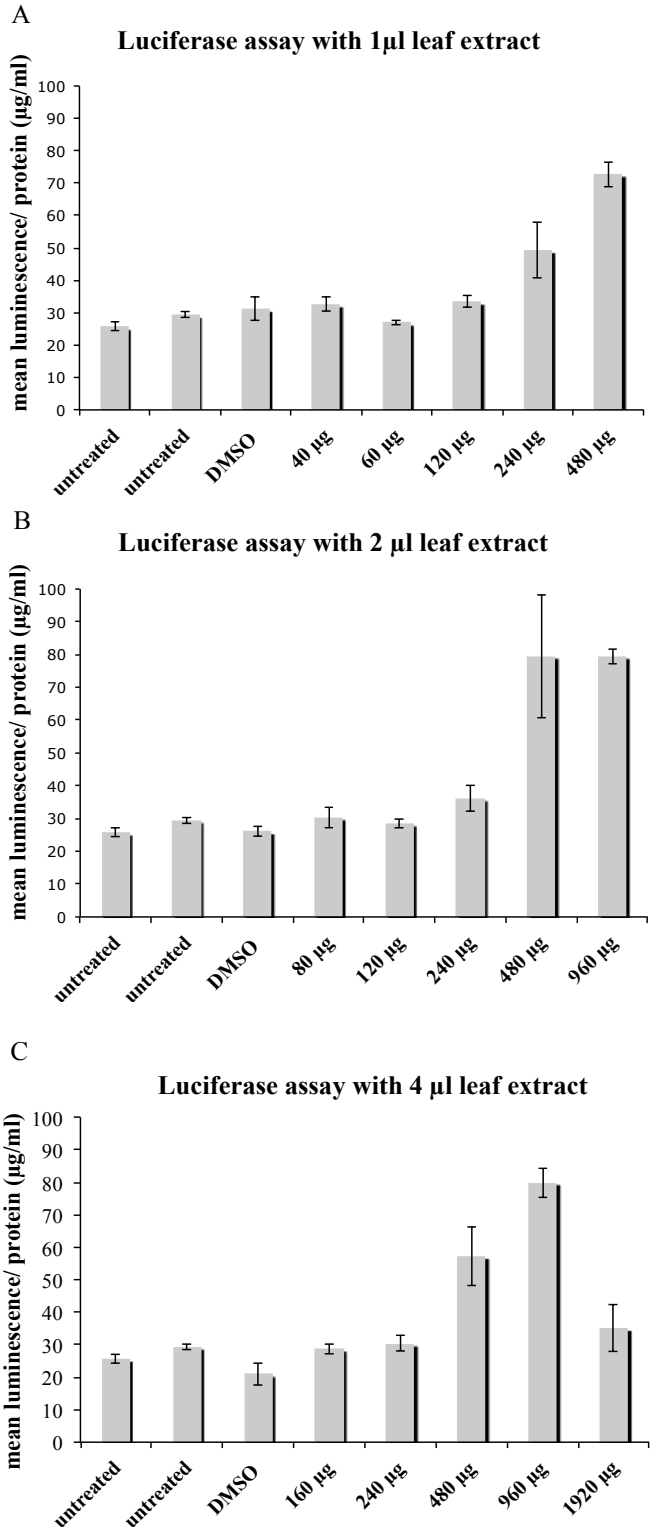


Fig. 16: Luciferase assay with different dilutions of leaf extract. Luminescence signal is given per µg/mL total cellular protein. (A) 1 µL of leaf extract/ DMSO in different dilutions with 40 µg, 60 µg, 120 µg and 240 µg of extract (B) 2 µL of leaf extract/ DMSO in different dilutions with 80 µg, 120 µg, 240 µg, 480 µg and 960 µg of extract (C) 4 µL of leaf extract/ DMSO in different dilutions with 160 µg, 240 µg, 480 µg, 960 µg and 1920 µg of extract. No extract was added to untreated cells; with DMSO treated cells were induced with 1, 2 or 4 µL DMSO. Bars represent the standard error of three technical replicates.

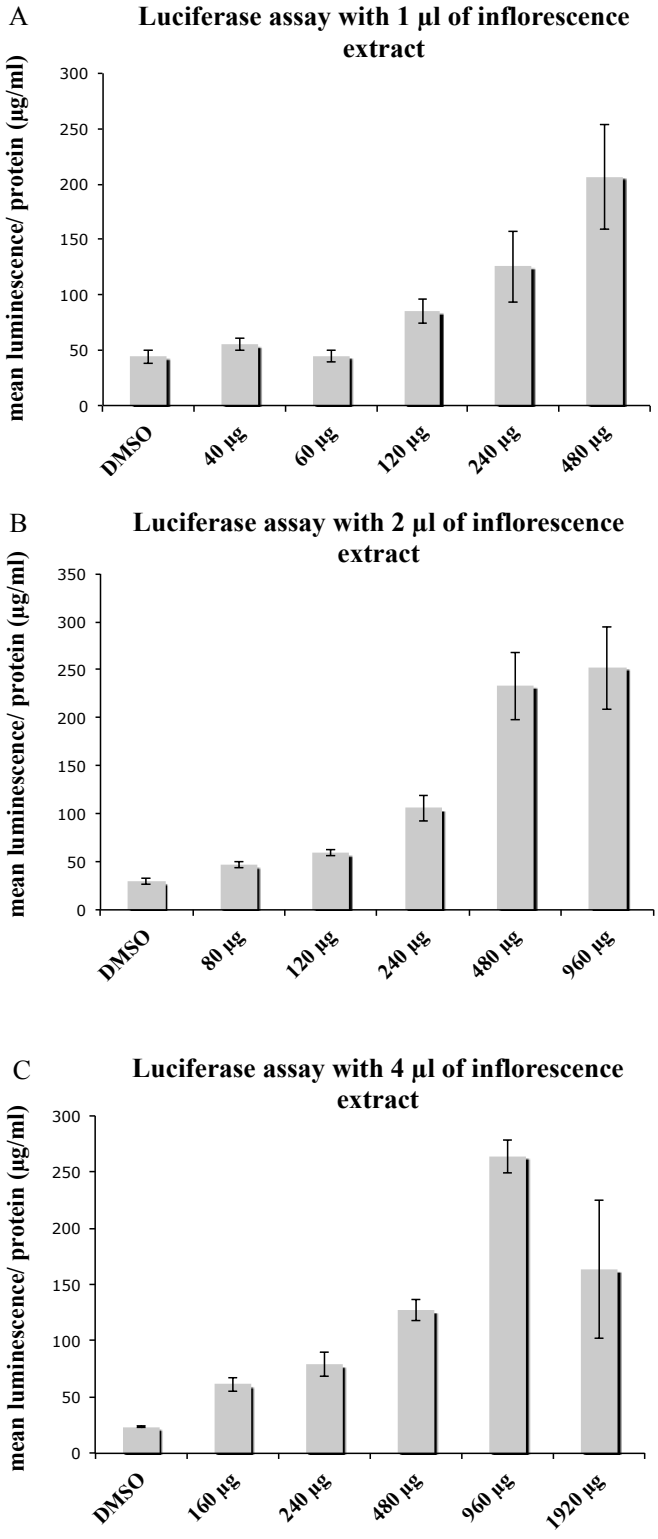


Fig. 17: Luciferase assay with different dilutions of inflorescence extract. Luminescence signal is given per µg/mL total cellular protein. (A) 1 µL of inflorescence extract/ DMSO in different dilutions with 40 µg, 60 µg, 120 µg and 240 µg of extract (B) 2 µL of inflorescence extract/ DMSO in different dilutions with 80 µg, 120 µg, 240 µg, 480 µg and 960 µg of extract (C) 4 µL of inflorescence extract/ DMSO in different dilutions with 160 µg, 240 µg, 480 µg, 960 µg and 1920 µg of extract. No extract was added to untreated cells; with DMSO treated cells were induced with 1, 2 or 4 µL DMSO. Bars represent the standard error of three technical replicates.

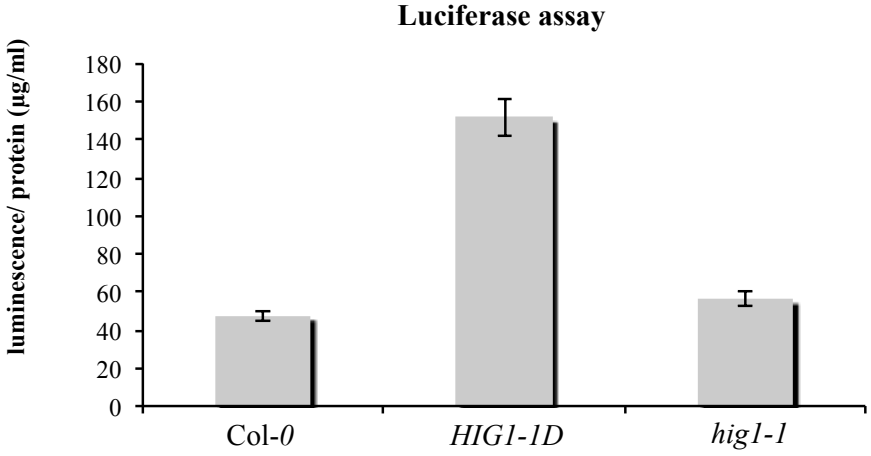


Fig. 18: Induction of Hepa1c1c7-Lux cells with extracts of mutants *HIG1-1D* and *hig1-1*. The bars represent the average of five technical replicates with standard error. The data were provided by T. Gigolashvili.

3.1.2 Screening results using Hepa1c1c7-Lux cells and mapping of mutants with a high induction potential

After a pre-selection of mutants in the HepG2-GFP system, approximately 470 mutants were screened using the Hepa1c1c7-Lux cells. In addition, mutants of the pools numbered with 96-98 were screened directly in the Hepa1c1c7-Lux cells without pre-screening. The mapping was done in parallel to the screening. A few mutants that were selected for mapping had a high induction potential in the HepG2-GFP system but this could later not be confirmed in the Hepa1c1c7-Lux cells. The screening results were evaluated by calculation of the ratio between the luminescence-signal induced by extracts of the mutant and of Col-*0*. These ratios are listed together with the mapping results in Table 10. When the chemoprotective effect of a mutant could be confirmed in the Hepa1c1c7-Lux cells, then this is noted with "yes" in the farthest right column of the table. Not confirmed mutants are marked with "no", "uncertain" means that the screening results from the HepG2-GFP cells could only partly be confirmed using Hepa1c1c7-Lux cells. This is the case when two or more biological replicates gave contradicting results. A ratio above "1", means that the mutant extract led to a higher induction than the extract of Col-*0*. Is the ratio below "1", the extract of the mutant did not induce the luminescence as strongly as Col-*0*, or

the extract may have had a toxic effect on the cells. Several ratios are listed for most of the mutants. These represent the results of different biological replicates, while one ratio is based on the average of three technical replicates. Since the luminescence signal of a screen is dependent on the quality of the substrate (age), the age of the cells (passage), the fitness and the age of the plants, the values of the screens cannot directly be compared with each other. Each screen was performed with a Col-*0* control. Therefore, the calculation of the ratios between mutant and Col-*0* induced signals leads to a comparable value. Some of the earlier mapped mutants could not be confirmed in the Hepa1c1c7-Lux cells. Since they were selected for mapping due to the induction of GFP in the HepG2-GFP cells, they were still considered for further analyses. The differing results may be due to the different backgrounds of the systems, human and mouse cells, and to the expression levels of the genes in the different TAMARA generations. In addition, putative toxic effects of high amounts of a compound may also lead to a false interpretation of screening results. All luciferase assay-screening results of the mapped mutants are listed in the supplementary. In addition the plant phenotype is shortly described. The results of the quinone reductase assay are also listed in the supplementary.

Table 10: Summary of the mapped mutants.

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
1	**39.45**	**At1g80760** **At1g80770**	boron transporter pigment defective 318 (ferrous iron transport)	18x 2x	1.6/ 0.71	uncertain
2	**39.46**	At2g39000 **At2g39010 (3. intron)** At2g39020	GCN5-related N-acetyltransferase (GNAT) family protein plasma membrane intrinsic protein; water channel activity (chloroplast) PIP2;6 GCN5-related N-acetyltransferase (GNAT) family protein (chloroplast)	0.75x 0.26x 0.65x	1.25/ 0.65	uncertain
3	**41.08**	At5g14830	transposable element		1.0	no
	41.77	**At5g14840 (1. exon)** At5g14850	pseudogene (unknown function) mannosyltransferase	ko 1x	n.d.	
4	**51.14**	At3g09240	protein kinase-related	17 x	0.48	no
	42.25	**At3g09250 (3. exon)**	DNA-binding/nuclease	0.32 x	1.80/ 0.65	uncertain
	66.71*	At3g09260	beta-glucosidase (PYK10)	730 x	1.05/ 1.16	yes
5	41.16	**At4g27250**	putative NAD-dependent epimerase/dehydratase	0.74 x	n.d.	n.d.
	41.66	**At4g27260**	IAA amydo synthase (AtGH3.5)	152 x	1.12/ 0.94/ 2.34	yes
	41.58°				1.55/ 1.20/ 2.13	yes
	40.36				1.64/ 0.58/ 1.97/ 2.09	yes
	40.76				1.12/ 0.80/ 0.64/ 1.06	uncertain
6	**42.63**	At1g25220 **At1g25230 (1. exon)** At1g25240	putative beta subunit of anthranilate synthase (AtASB1) putative calcineurin-like phosphoesterase putative AP180-like endocytotic accessory protein	1x 45x 9x	1.44/ 1.36	yes

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
7	43.40	At2g16340	unknown protein	130x	1.20/ 1.52	yes
	43.09	**At2g16365 6. (intron)**	F-box family protein	80x	1.91/ 1.67	yes
	49.07	At2g16360	S25-type protein of small ribosomal subunit (AtRPS25A)		n.d.	n.d.
	65.20				0.60	no
8	**43.62**	At4g39640	gamma-glutamyl transpeptidase (AtGGT1)	1.4x	1.62/ 0.43	uncertain
		At4g39650 (1. intron)	gamma-glutamyl transpeptidase (AtGGT2)	1.22x		
		At4g39660	alanine:glyoxylate aminotransferase (AtAGT2)	0.84x		
9	**47.15**	At5g15490	UDP-glucose 6-dehydrogenase, putative (AtUDG3)		2.09/ 0.32	uncertain
		At5g15500 (1. exon)	ankyrin repeat family protein/putative membrane protein of unknown function	4.6x		
		At5g15510	protein of unknown function, TPX2 (targeting protein for Xklp2 (mitotic spindle assembly))	124x		
10	**47.25°**	**At5g66570**	membrane-extrinsic subunit of photosystem II complex (AtPsbO1)	7x	0.95	no
		At5g66580	unknown protein	3.5x		
11	**47.64**	**At4g37290**	unknown protein	22x	1.73/ 0.69	uncertain
		At4g37295	unknown protein	120x		
12	49.34	**At2g45660 (5'UTR)**	MADS-box-type transcription factor (AtSOC1/AtAGL20)	1.74x	0.82	no
	67.31	**At2g45670**	lysophosphatidylethanolamine acyltransferase (AtLPEAT2)	2.36x	1.23/ 1.29	yes
13	**49.44**	At4g38990	putative glycosyl hydrolase (AtGH9B16)	48.6x	1.26/ 1.89	yes
	49.41	**At4g39000 (2. exon)**	putative endo-1,4-beta glucanase	7.3x	n.d.	
		At4g39010	putative endo-1,4-beta glucanase	0.53x		

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
14	49.62	2 Insertions:			1.58/ 0.71	uncertain
	49.53	**At5g13330**	putative ERF-type transcription factor	9.5x	1.17	yes
	49.72°	**At5g13340**	unknown protein	12.3x	n.d.	n.d.
	49.69°				n.d.	n.d.
	49.64	At5g52200	putative protein phosphatase inhibitor 2 (IPP-2)	9x	n.d.	n.d.
	49.59	**At1g52190 (3.exon)**	putative nitrate transporter	ko	0.71	no
	51.34	At1g52185	micro RNA	2x	0.47	no
	51.23	At1g52180	putative MIP-type aquaporin or small neutral solute transporter	36x	n.d.	n.d.
15	**50.02°**	At2g38280	membrane-associated AMP deaminase, embryonic factor (AtFAC1	12x	0.60	no
		At2g38290	ammonium transporter (AtAMT2.1)	14x		
		At2g38300	putative GARP-G2-type transcription factor	0.75x		
16	**50.36**	**At1g04360**	putative ubiquitin ligase, ATL subfamily (AtATL1)	50x	n.d.	n.d.
		At1g04370	ethylene-responsive element binding factor (AtERF14)	13000x		
		At1g04380	putative 2OG-Fe(II) oxygenase	50x		
17	50.32	**At2g34830 (1.exon)**	putative WRKY-type transcription factor (AtWRKY35)	740x	1.10/ 0.45	no
	50.54	At2g34825	RALF-type endogenous peptide (AtRALFL20)	100x	0.41	no
	49.59	At2g34820	putative bHLH-type transcription factor	50x	0.71	no
	50.52				0.41	no
18	**50.46**	**At2g43610**	putative chitinase	120x	1.00	no
	50.56	**At2g43620**	putative chitinase	100x	0.72	no
19	**51.70**	**At1g01460**	putative phosphatidylinositol phosphate 5-kinase (AtPIP5K10)	70x	0.55	no
		At1g01470	putative LEA protein (LEA14)	1.4x		
20	**52.18**	**At3g54840**	putative RAB-F-class small GTPase (AtRAB-F1/AtARA6)	1x	0.94	no
	46.24	**At3g54830**	putative Amino acid transporter	80x	2.90/ 0.81	uncertain
	46.23				0.64	no

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
21	**52.38**	At1g14410	Whirly-type transcription factor (AtWHY1)	5x	n.d.	n.d.
		At1g14420	putative pectate lyase	7x		
		At1g14430	putative glyoxal oxidase	13x		
22	**53.11**	At5g43180	putative membrane protein of unknown function	35x	0.63	no
		At5g43185 (1. exon)	unknown protein	0.01x		
		At5g43190	F-box family protein (FBX6); ubiquitin-protein ligase activity	2x		
		At3g63050	unknown protein	31x		
		At3g63040	unknown protein	220000x		
		At3g63030	protein containing methyl-CpG-binding domain. Has sequence similarity to human MBD proteins	3x		
23	**55.16**	At1g31880	BRX-type transcription factor, controls extent of cell proliferation and elongation (AtBRX)	12x	0.63	no
		At1g31885 (3. exon)	NOD26-like intrinsic protein (AtNIP3.1/AtNLM9)	87x		
		At1g31910	putative phosphomevalonate kinase	3x		
24	**55.53**	At3g57680	putative protein C-terminal processing serine protease	1x	0.76	no
	55.26	**At3g57670 (2. exon)**	putative subclass A1d C2H2-zinc-finger-type nucleic acid binding protein (AtA1d-03/AtWIP2)	91x	n.d.	n.d.
	55.38	At3g57660	putative 1st subunit of DNA-directed RNA polymerase	1x	0.83	no
	48.69				1.91/ 1.21	yes
25	**59.67**	**At1g05240/At1g05250 are 100% homolog (2. intron)**	putative class-III peroxidase (AtPer1)/(AtPer2)	180x	n.d.	n.d.
		At1g05260	putative class-III peroxidase (AtPer3), RCI3, cold-inducible cationic peroxidase	25000x		
26	**61.64**	At3g07850	putative galacturonase/polygalacturonase	50x	n.d.	n.d.
	61.52*	At3g07860	unknown protein	0.65x	0.55/ 1.66	uncertain
	61.63*	**At3g07870 (1.exon)**	protein of unknown function, contains F-box domain	0.03x	0.43/ 1.04	uncertain
	75.65*				0.89/ 0.33	uncertain
	75.67*				6.76	yes
	76.65*				5.59	yes

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
27	**61.36**	At3g11130 **At3g11150** **At3g11160**	putative heavy chain of clathrin complex (AtCHC1) unknown protein unknown protein	1.4x 86x 135x	n.d.	n.d.
28	**62.40**	At3g60640 At3g60647 **At3g60650 (1. exon)** At3g60660	ubiquitin-like protein, involved in protein degradation (AtATG8g), AUTOPHAGY 8G unknown protein unknown protein unknown protein	2.6x 286000x 3500x 12x	n.d.	n.d.
29	**63.16**	**At1g08670** **At1g08680**	epsin N-terminal homology (ENTH) domain-containing protein / clathrin assembly protein-related ADP ribosylation factor GTPase-activating protein (AtAGD14/AtZIGA4)	1596x 1.5x	2.62/ 1.55	yes
30	**63.17**	At2g19620 **At2g19610 (2. exon)** At2g19600	putative Ndr-type hydrolase C3HC4-type RING finger family protein putative potassium cation effluc antiporter (AtKEA4)	0.8x 38x 2x	2.83/ 2.46	yes
31	**63.67**	At5g65683 **At5g65690 (1.intron)** At5g65687	zinc finger (C3HC4-type RING finger) family protein phosphoenolpyruvate carboxykinase carbohydrate transmembrane transporter activity	1.6x 7300x 2x	1.70/ 0.90	uncertain
32	**64.15**	**At5g11530** **At5g11520**	embryonic flower 1 (putative transcription factor, involved in regulating vegetative and reproductive growth) aspartate aminotransferase (AtASP3)	0.28x 0.036x	2.02/ 1.52	yes
33	**64.02**	At1g07190 **At1g07180 (1. intron)** At1g07175	unknown protein putative alternative NADPH/NADH-dependent dehydrogenase (AtNDA1) putative signaling tyrosine-sulfated glycopeptide (18 aa) precursor protein	0.0579x 20x	1.28/ 2.07	yes
34	**64.22**	**At2g19893** **At2g19890**	defensin-like (DEFL) family protein protein of unknown function	150000x 31916x	1.00/ 2.67	uncertain

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
35	**64.47**	At1g20490	putative acyl-CoA synthetase (AtACS2,) AMP-dependent synthetase, ligase	2x	2.68/ 1.79	yes
		At1g20500 (1. exon)	putative acyl-CoA synthetase (AtACS3), AMP-dependent synthetase, ligase	37 000x		
		At1g20510	putative acyl-CoA synthetase (AtACS4/AtOPCL1)	8.6x		
36	**64.77**	**At3g60966**	putative C3HC4-type zinc finger ubiquitin E3 ligase	0.05x	1.73/ 0.77/ 0.90	no
		At3g60970	putative subfamily C ABC-type transporter (AtABCC15/AtMRP15)	0.18x		
37	**64.48**	**At1g70970**	protein of unknown function, contains F-box domain	2670x	3.05/ 1.91	yes
		At1g70980	putative asparaginyl-tRNA synthetase (AtSYNC3)	34x		
38	**64.59**	**At2g38300**	putative GARP-G2-type transcription factor	3.6x	2.79/ 2.03	yes
	65.28	**At2g38290**	ammonium transporter (AtAMT2.1)	7.7x	1.29/ 0.84	uncertain
	72.68				1.29/ 1.78	yes
39	**65.32**	**At1g63810**	putative Nrap-like nucleolar RNA-associated protein	0.79x	1.27/ 1.95	yes
	68.71*	**At1g63820**	protein of unknown function, contains CCT domain	46.7x	0.70/ 3.56	uncertain
	67.18				1.02/ 1.07	uncertain
40	**67.51**	**At1g47813**	protein of unknown function	0.21x	1.32/ 0.77	uncertain
		At1g47810	protein of unknown function, contains F-box domain	154x		
41	**67.74**	At3g04560	unknown protein	0.63x	1.77/ 1.29	yes
		At3g04570 (1. exon)	protein of unknown function, contains AT-hook-type DNA binding motif	5.7x		
		At3g04580	putative ethylene-responsive sensor histidine protein kinase (AtEIN4)	1.2x		

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
42	**68.15***	**At2g37440**	inositol polyphosphate 5-phosphatase	0.05x	2.92	yes
		At2g37450	unknown function (Plant Drug/Metabolite Exporter)	1.15x		
		At2g38290	Ammonium transporter	1.06x		
		At2g38300	putative GARP-G2-type transcription factor	0.88x		
	81.77*	**At2g37440**	inositol polyphosphate 5-phosphatase	1790x		
		At2g37450	unknown protein (Plant Drug/Metabolite Exporter)	2x		
43	**68.24***	**At1g71002**	miRNA (targets MYB family members)	342(N)x/1320(C)x	1.04/ 2.00	yes
		At1g71000	Hsp40/DnaJ-type molecular chaperone	4.5x		
44	**68.41***	At5g41730	protein kinase family		2.54	yes
		At5g41740 (3. exon)	TIR-NBS-LRR class disease resistance protein			
		At5g41750	TIR-NBS-LRR class disease resistance protein			
45	**70.14**	At5g22510	putative beta fructofuranosidase/invertase/saccharase	1.7x	1.17/ 1.00	uncertain
	50.66	**At5g22500 (1.intron)**	fatty acid reductase 1	4x	1.34/ 1.54	yes
	88.36*	At5g22490	condensation domain containing protein	34.7x	4.02	yes
46	**70.24**	**At5g59340**	homeobox-type transcription factor WUSCHEL RELATED HOMEOBOX 2 (WOX2)	100x	1.12/ 0.89	uncertain
	71.71	**At5g59330**	lipid binding (involved lipid transport, in endomembrane system)	550x	1.58/ 0.91	uncertain
	57.31				0.68	no
	73.30*				0.53/ 1.01	no
	70.06				2.44/ 1.06	yes

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
47	**70.70**	At5g49120 **At5g49130 (1.exon)** At5g49140	protein of unknown function putative MATE-related effluc carrier (AtDTX55) disease resistance protein (TIR-NBS-LRR class), putative	32x 100x 1300x	1.15/ 1.63	yes
48	**71.12**	At2g30200 **At2g30210 (4. exon)** At2g30220	putative malonyl-CoA:ACP transacylase putative laccase (AtLAC3) putative GDSL-type lipase	1.6x 547x 57.1x	1.97/ 2.24	yes
49	**71.34**	At3g03470 **At3g03480 (1.exon)** At3g03490	cytochrome P450 monooxygenase (AtCYP89A9) putative BAHD-type acyltransferase putative peroxisomal biogenesis factor (AtPEX19-1)	1.52x 1.74x 2.11x	1.28/ 2.30	yes
50	**71.36**	**At3g04100** **At3g04110**	MADS-box-type transcription factor (AtAGL57) putative GLR-type calcium cation-permeable channel (AtGLR1.1)	42x 0.17x	2.58/ 1.55	yes
51	**71.72**	**At5g67411**	putative GRAS-type transcription factor	1.3x	2.68/ 0.83	uncertain
	63.01	**At5g67420**	putative ASL/LBD-type transcription factor (AtASL39/AtLBD37)	1x	2.07/ 2.32	yes
52	**72.14**	At1g02950	putative class phi glutathione S-transferase (AtGSTF4)	1.82x	1.13/ 1.84	yes
	72.08	**At1g02940 (1. intron)** At1g02930	putative class phi glutathione S-transferase (AtGSTF5) putative class phi glutathione S-transferase (AtGSTF6)	130x 56x	1.31/ 1.70	yes
53	**72.29**	At3g59580 **At3g59590 (3. exon)** At3g59600 At5g59845 **At5g59840 (4. exon)** At5g595830 At2g23130 **At2g23140 (3. intron)** At2g23142	putative Nin-like-type transcription factor putative jacalin-type lectin putative 8th subunit of DNA-directed RNA polymerase putative GASA/GAST/Snakin-type gibberellin-regulated protein putative RAB-E-class small GTPase (AtRAB-E1b) unknown protein putative lysine-rich arabinogalactan protein (AtAGP17K) putative U-box-type E3 ubiquitin ligase (AtPUB4) putative S1 self-incompatibility protein	1.19x 0.04 1.6x	1.28/ 2.44	yes

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
54	**72.67**	At4g37580	GNAT-type N-acetyltransferase. involved in apical hook development (AtHLS1)	2x	1.15/ 2.81	yes
		At4g37570 (1. intron)	intron of transposable element gene	1.5x		
		At4g37560	formamidase, putative / formamide amidohydrolase	1.6x		
55	**73.63***	At1g05650	putative galacturonase/polygalacturonase	126200x		
		At1g05660 (4. exon)	putative galacturonase/polygalacturonase	314800x	0.65/ 1.05	no
		At1g05670	UDP-dependent glycosyl transferase (AtUGT74E1)	878x		
56	**75.66***	At3g13890	putative Myb-type transcription factor (AtMYB26)	26.9x	5.53	yes
		At3g13900 (10. exon)	aminophospholipid translocase (P4-type ATPase (AtALA7))	0.81x		
		At3g13910	unknown	0.75x		
57	**79.23***	At3g27500	protein of unknown function, contains C1-type domain	15.2x	1.01/ 2.31	uncertain
		At3g27503	putative small, secreted, cysteine rich protein (AtSCRL19)	28x		
		At3g27510 (1. exon)	protein of unknown function, contains C1-type domain	0x		
		At3g27520	unknown protein	1.8x		
58	**81.21***	At5g46530	AWPM-19-like membrane protein of unknown function	1.67x	0.90/ 0.72	no
		At5g46540 (8. exon)	putative subfamily B ABC-type transporter (AtABCB7/AtMDR7)	8213x		
		At5g46550	unknown protein (contains bromodomain)	1.22x		
59	**82.20***	**At3g01070**	ENOD-like protein (AtENODL16)	3.6x	1.19/ 0.72	uncertain
		At3g01080	WRKY-type transcription factor, WRKY58	1.7x		
60	**84.20***	**At5g65630**	unknown protein (contains bromodomain)	1x	1.05/ 1.62	yes
	84.27*	**At5g65640**	bHLH-type transcription factor bHLH093	7x	2.76/ 7.97	yes
61	**84.63***	At1g61260	unknown protein		1.85/ 1.64	yes
		At1g61270 (3. exon)	lysine/histidine transporter (AtLHT3) (Amino Acid/Auxin Permease)			
		At1g61275	snRNA	.		

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
62	**84.64***	**At5g59730**	exo70-type subunit of Exocyst complex (AtExo70H7)	5.12x	0.45/ 0.99	no
		At5g59740	UDP-galactose/UDP-glucose transporter (UDP-Galactose:UMP Antiporter)	4.58x		
63	**84.75***	**At4g30350**	ATP-dependent Clp-type protease	2.8x	0.38/ 4.09	uncertain
		At4g30360	cyclic nucleotide and calmodulin-regulated ion channel (AtCNGC17)	0.7x		
64	**86.25***	**At5g46980**	invertase/pectin methylesterase inhibitor	57x	8.47/ 1.97	yes
		AT5g46990	invertase/pectin methylesterase inhibitor	9x		
		At3g63530	E3 ubiquitin ligase (BIG BROTHER)	0.4x		
		At3g63540 (1. exon)	unknown protein	1.4x		
65	**86.56***	**At3g02000**	putative glutaredoxin (AtROXY1)		3.04	yes
		At3g02010	protein of unknown function, contains pentatricopeptide (PPR) repeat			
66	**86.71***	**At5g11320**	flavin-containing monooxygenase (AtYUC4)		1.82	yes
		At5g11330	putative monooxygenase	1x		
67	**86.76***	**At5g28370**	unknown protein (contains pentatricopeptide (PPR) repeat)	1.23x	2.98/ 2.75	yes
		At5g28380	unknown protein (contains pentatricopeptide (PPR) repeat)	0.25x		
68	**87.09***	**At1g62886**	putative (yeast TFB5)-like subunit of TFIIH basal transcription factor complex	3.46	2.13	yes
		At1g62880	putative membrane protein of unknown function/cornichon protein	3.38		
69	**87.17***	**At1g30840 (1. exon)**	purine permease (AtPUP10)	0.9x	1.48	yes
		At1g30845	putative membrane protein of unknown function	2.06x		
		At1g28300	putative LAV family B3-type transcription factor (AtLEC2)	35.6x		
		At1g28307	unknown protein	10.3x		
		At1g28310	putative Dof-type zinc finger transcription factor	1.14x		

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
70	**87.42***	**At1g15757 (2. exon)** At1g15760	putative defensin-like protein unknown protein	4.53x 179.4x	3.75	yes
71	**87.58***	**At5g10180 (5' UTR)** At5g10190	low-affinity sulfate transporter (AtSultr2.1) (Sulfate Permease) small solute transporter	1.013x 1690x	2.98	yes
72	**87.59***	At2g41850 **At2g41860 (2. exon)** At2g41870	polygalacturonase (AtADPG2/AtPGAZAT) calcium-dependent protein kinase (AtCPK14) putative remorin (AtREM4.2)	71x 100570x 3.27x	2.24	yes
73	**88.05***	At5g35760 **At5g35770 (2.exon)**	Beta-galactosidase related protein Sterile Apetala protein (AtSAP) (TF activity)	26.6x 674x	1.43	yes
74	88.06*	**At5g55880**	unknown protein	100088x	2.24	yes
	89.29*	**At5g55890**	unknown protein		3.23	yes
	89.37*				1.56	yes
	90.32*				3.05	yes
	90.40*				4.43	yes
	90.43*				3.56	yes
	90.53*				2.12	yes
75	**88.75***	**At1g52210** **At1g52220**	transposable element gene subunit PsaP of photosystem I complex		3.25	yes
76	**89.68***	**At1g79110** **At1g79120** At3g12900 **At3g12910 (3. exon)** At3g12915	protein of unknown function, contains RING-type zinc finger domain putative ubiquitin carboxyl-terminal hydrolase putative 2OG-Fe(II) oxygenase putative NAC/NAM-type transcription factor putative type 2 translation elongation factor	1.84x 1.77x 2.6x 1.84x 0.56x	2.79	yes
77	**90.63***	At1g77180 **At1g77200 (1. exon)** At1g77210	putative TF response to abscisic acid, salt and osmotic stress putative ERF-type transcription factor STP subfamily monosaccharide:proton symporter (AtSTP14)	1.36x 0.42x 0.93x	3.09	yes

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
78	**90.36***	At3g10113	putative MYB_related-type transcription factor	1.85x	2.66	yes
	90.37*	**At3g10114**	pseudogene of glycosyl hydrolase family 81 protein (unknown function)	8.41x	2.67	yes
		At3g10116	putative GPI-anchored COBRA-type protein of unknown function	5.08x		
		At1g01600	Cyp86A4 subfamily of cytochrome p450 genes	321x		
		At1g01610 (3. intron)	protein with glycerol-3-phosphate acyltransferasse activity	11.5x		
		At1g01620	plasma membrane intrinsic protein	0.17x		
79	**90.46***	At5g09520	putative proline-rich glycoprotein (AtPRP9)	67.7x	2.17	yes
		At5g09530 (1. exon)	putative proline-rich glycoprotein (AtPRP10)	16.7x		
		At5g09540	putative DnaJ-type molecular chaperone	88.7x		
80	**90.54***	**At3g15500**	NAC-type transcription factor (AtANAC055/AtNAC3)	2.33x	3.18	yes
		At3g15510	NAC-type transcription factor (AtNAC2)	4.7x		
81	**90.64***	At2g45040	putative matrixin-type metalloprotease (AtMMP4)	75.7x	4.11/ 4.27/ 2.54	yes
		At2g45050 (3' UTR)	putative GATA-type transcription factor (AtGATA-2)	153.8x		
		At2g45060	unknown protein	1.89x		
82	**94.31***	At4g40000	NOL1/NOP2/sun-type protein	41.7x	0.64	no
	94.32	**At4g40010 (6. intron)**	SNF1-related protein kinase, SnRK2 family (AtSnRK2.7/AtSRK2F/AtOSKL5)	114.5x	1.21	yes
	94.10	At4g40011	unknown protein	4.01x	n.d.	n.d.
83	**94.12***	**At3g63540 (1. exon)**	unknown protein	0.023x	0.58/ 0.86/ 0.45	no
	94.52*	At3g63530	E3 ubiquitin ligase (BIG BROTHER)	3.6x	0.76	no
	94.67*				0.75	no
	94.49*				0.48	no

Nr.	Mutant	Insertion site/ flanking genes	Description of genes	Relative gene expression	Ratio Luc	Confirmed with Luc
84	**94.48***	**At1g75440** **At1g75450**	putative ubiquitin-conjugating E2 enzyme (AtUBC16) cytokinin oxidase (AtCKX5)	129x 167x	0.45	no
85	**95.16**	At4g11910	senescence-associated protein (regulates chlorophyll degradation)	1.9x	1.98	yes
	95.18	At4g11911	putative senescence-associated protein, probably involved in regulation of chlorophyll degradation	69x	1.52	yes
	95.22	At4g11920	putative Cdc20-like mitotic specificity factor for anaphase-promoting complex	30x	2.07	yes
	95.31				1.89	yes
86	**95.19**	At3g01860 At3g01870 At3g01880	unknown protein unknown protein unknown protein	10.72x 52x 1.15x	1.73	yes
87	**95.36**	At1g11915 At1g11920 At1g11925	unknown protein pectate lyase family protein Stig1 family protein	0.56x 2.38x 0.39x	3.55	yes

Note: Part of the mapping was performed by Eigen, C. (*167*), and Krizowski, S (*165*). The data from these screenings are marked with an asterisk (*) for Eigen, C. and a circle (°) for Krizowski, S. The relative gene expression is given for those mutants written in bold. n.d.= not determined, ko = knockout. yes = mutant confirmed, uncertain = contradicting results, no = mutant not confirmed. The ratio is calculated by dividing the luminescence value of the mutant by the luminescence value of Col-*0*.

3.1.3 Mapping and expression analysis of the chemoprotective mutants

After the mutants had been screened, they were further analyzed by mapping the insertion site of the T-DNA (2.4.4). The expression levels of the flanking genes were determined via real time PCR analysis (2.7.1). From the approximately 5,000 mutants originally screened, 149 mutants were mapped (3%) (see Fig. 19).

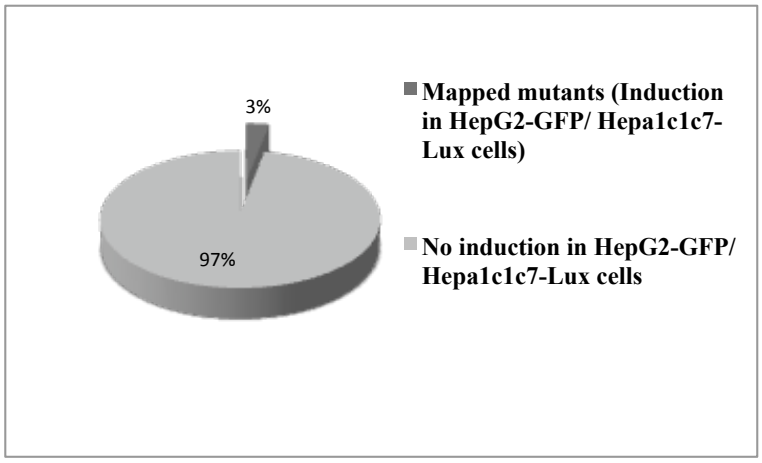


Fig. 19: Pie diagram of the percent of mutants chosen for mapping from the total number of screened mutants.

The mapping uncovered that some mutants have the same insertion site, suggesting that they are derived from the same parent plant. Therefore, after comparison of all insertion sites, 87 mutant lines with different insertion sites were counted. 29% of the detected insertion sites were identified in at least two different independent mappings of selected mutants (see Fig. 20).

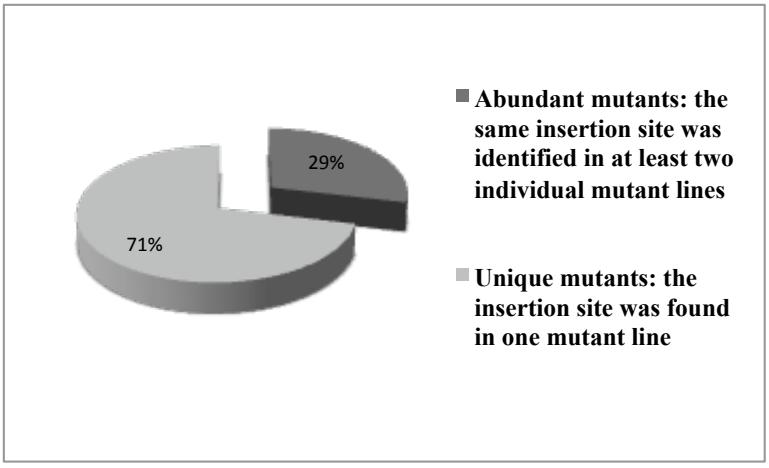


Fig. 20: Pie diagram of the percent of insertion sites found in at least two different mutant lines.

Table 10 summarizes the mutants that were mapped. The table includes the gene At-number, a short description of the putative or known gene function as annotated in the online databases ARAMEMNON (*159*) or TAIR (*158*) and the relative expression levels of the genes flanking the insertion site for one of the mutants listed (marked in bold letters). The expression levels listed are designed to give an orientation of how strongly a gene is influenced. Expression analyses of TAMARA lines have shown that the expression level is not stable in all lines (data not shown).

Those genes that show an altered expression in comparison to wild-type or that contain the T-DNA insertion and are therefore most likely affected in their functionality can be divided into different groups or clusters based on their putative or known gene function and annotation. In summary, 158 genes were categorized. 32% of these genes encode for (putative) enzymes, 13% (putative) transcription factors, 11% of the genes encode for (putative) transporters or are involved in transport, 26% of the genes encode for (putative) proteins of unknown function and 18% belong to other categories. Among this category are many proteins with functions in housekeeping/maintenance of life and cell structure. The latter group is not categorized because some of the representatives in this group can also be sorted to other categories and vice versa. Examples are a number of enzymes such as an asparaginyl-tRNA synthetase, an ubiquitin-conjugating E2 enzyme, a protein kinase-related enzyme, a nuclease, and an ATP-dependent Clp-type protease. Genes involved in plant defense and/or stress response can be detected in several categories. Examples are putative chitinases in the enzyme category, ethylene responsive transcription factors or defensin like family protein in the mixed gene group. The latter group also contains inhibitors, pseudogenes, subunits of protein complexes, glycoproteins and proteins involved in the cell cycle such as the putative Cdc20-like mitotic specificity factor for anaphase-promoting complex or the protein TPX2 involved in mitotic spindle assembly.

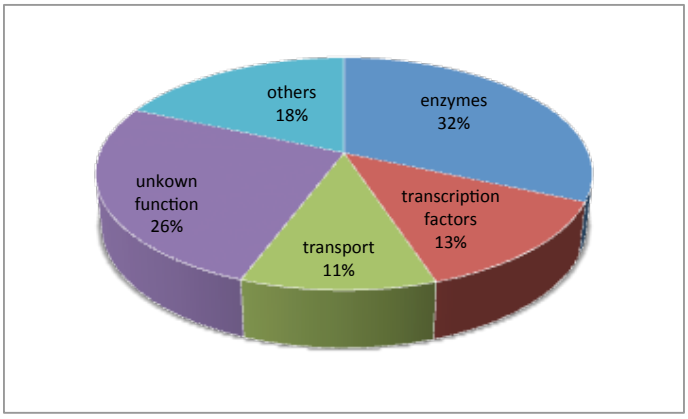


Fig. 21: Pie diagram of the functional gene categorization. All genes that showed an altered expression in the chemoprotective mutants are included.

More precise analyses can be performed using the online bioinformatics tool DAVID (*168*). With this tool, the genes can, among others, be classified into functionally related gene groups (Gene Functional Classification) or on the basis of functional annotation (Functional Annotation Clustering). Gene functional classification summarizes the major roles of the discovered gene groups by generating a gene-to-gene similarity matrix. This is based on the functional annotation from different sources. Functional Annotation Clustering is based on the theory that similar annotations have similar gene members, so that genes are grouped depending on the amount of shared gene annotations. The gene functional classification (medium classification stringency, default setup) of the 158 genes resulted in six clusters, however a high number of genes are excluded from the clustering due to unknown functions or categories with only one gene member (data not shown).

Functional annotation clustering analysis results in 25 clusters (default settings, medium classification stringency). The results of the functional annotation clustering are shown in Table 11. The members of each of the 25 clusters are listed as well as the features of the members of a cluster. Not all gene members of a cluster have all of the described features; therefore the detailed data are shown in the supplementary material. Many of the genes are categorized into more than one group making some of the clusters similar to each other and consisting of almost the same genes. Twelve genes form the cluster I. In this cluster members exhibit glycosidase activities and are involved in cell wall organization. Some of the genes encode for proteins closely related to pectin lyase, which act on forms of pectin found in the cell wall. Cluster II

consists of 14 members. Ten members of this group are involved in transport. Among this group are several aquaporins. The main characteristic of cluster III is hydrolase activity, followed by glycosidase activity, involvement in cell wall organization, polysaccharide and carbohydrate metabolism. Part of this group, are also three members of the glycoside hydrolase family 9. Most members of cluster IV contain a signal peptide and/or are found in the extracellular region and are glycosylated. All eight genes of cluster V are involved in reproductive structure development. The members of cluster IX are identical with those of cluster V but for two further genes found in cluster IX. The additional gene terms describing cluster IX include the involvement in the reproductive developmental process and postembryonic development. Furthermore, the members of this cluster can be divided into a group of genes involved in flower development and a group with functions in seed and fruit development. Cluster VI consists of genes encoding for proteins that bind to calcium ions and/or have an active site as a proton acceptor. Most members of cluster VII bind to DNA and/or have transcription factor activity, and/or are involved in transcription regulation. All genes of cluster VIII are integral/intrinsic to membranes. Some are transmembrane proteins and involved in transport. The genes of cluster X are involved in metabolic processes. Five out of six genes have functions in secondary metabolism, four in cellular amino acid derivative metabolic processes and three of the genes encode for proteins of the phenylpropanoid pathway. All 26 members of cluster XI bind to metal ions. Nearly half of the genes encode for proteins that bind to zinc ions. The members of clusters XII, XV and XX are identical. The three clusters have in common, that they respond to abiotic stimuli. Cluster XII highlights the members that respond to water deprivation, cluster XV those genes that respond to osmotic and salt stress and cluster XX the genes that respond to radiation/light. Most genes of cluster XIII are involved in intracellular signaling cascades; this can include responding to a hormone stimulus, to endogenous stimuli and organic substances. Three of the ten members respond to ethylene. The three genes forming cluster XIV all exhibit ATPase activity and are involved in the transmembrane transport of substances. Cluster XVI contains members that have zinc finger domains. All ten proteins bind to zinc ions. Seven of nine members of cluster XVII play a role in oxidation and reduction processes. Some of the proteins bind to iron ions. Cluster XVIII consists of four genes that all contain a F-box associated

domain type I. Three of the members are annotated to contain an F-box domain. The genes of cluster XIX encode for proteins with roles in macromolecule catabolism and/or proteolysis. Three of the proteins are hypothesized to be involved in ubiquitination pathways. The 17 members of cluster XXI all bind to purine nucleotides. Most of the proteins also bind to purine ribonucleotides and to ATP. Cluster XXII consists of genes that encode for proteins with disulfide bonds. Three out of five respond to fungi. The three members of cluster XXIII are localized in an organelle and the members of cluster XXIV, of which two are identical with cluster XXIV all have a transit peptide and are targeted to chloroplasts. Cluster XXV consists of three protein kinases.

In summary, considering that many clusters have common members, the functional annotation clustering divided the genes into the following major functional groups: Enzymes involved in macromolecule catabolic processes, such as polygalacturonases and glycosidases, playing a role in cell wall organization. Many of the enzymes are secreted. Some of the genes respond to abiotic or biotic factors, such as water, light and fungi. The second category of genes encode for proteins involved in transport. These proteins have transmembrane domains. Both passive and active transporters were found. The third group consists of genes involved in reproductive organ development such as the floral organs and the seeds. This group also contains proteins with transcription factor activity. Cluster X does not overlap with any other cluster. Its members are involved in secondary metabolic processes, such as the phenylpropanoid metabolism. The fifth group consists of genes that respond to abiotic stresses. Another group of genes that also contains transcription factors responds to hormones such as ethylene. There is also a group of proteins that bind to ions, such as proteins with a zinc finger domain, and finally a group of F-box proteins.

Table 11: Summary of the 25 clusters formed of the detected genes on the basis of functional annotation analysis.

Cluster	Characteristics	Genes
I	polygalacturonase; glycosidase; Parallel beta-helix repeat; polygalacturonase activity; Pectin lyase fold; cell wall organization	At1g14420; At1g11920; At2g43620; At3g09260; At2g43610; At4g39000; At4g39010; At4g38990; At3g07850; At1g05660; At2g41850; At1g05650
II	aquaporin; transport; topological domain: Extracellular; topological domain: Cytoplasmic; plasma membrane; short sequence motif:NPA 1; Major intrinsic protein; water channel activity; substrate specific channel activity; water transporter activity; passive transmembrane transporter activity	At2g39010; At1g01620; At1g80760; At1g52180; At3g04110; At1g31880; At1g31885; At2g38290; At3g13900; At3g60970; At5g59740; At5g46540; At2g43610; At4g11920
III	hydrolase; glycosidase; cell wall organization; macromolecule catabolic process; carbohydrate metabolism; polysaccharide metabolic process; cellulose degradation; cellulose catabolic process; glucan catabolic process	At4g39000; At4g39010; At4g38990; At2g43610; At2g43620; At3g07850; At1g05660; At1g05650; At2g41850; At3g09260; At2g45040; At3g60970; At3g13900; At2g38280; At2g30220; At1g25230; At1g04360; At4g11920; At1g75440
IV	Secreted; glycoprotein; extracellular region; signal peptide	At2g34825; At4g39650; At5g66570; At2g43620; At2g45040; At5g46540; At1g11920; At1g14420; At3g04110; At3g09260; At2g41850; At2g19893; At3g27503; At1g05240; At1g75450; At4g38990; At1g05260; At2g30220; At4g39000; At1g25230; At2g30210; At4g39010; At3g07850
V	multicellular organism reproduction; reproductive process in a multicellular organism; reproductive structure development	At5g59340; At2g38280; At2g34830; At3g63530; At5g35770; At1g28300; At2g41850; At3g13890
VI	active site: Proton acceptor; calcium ion binding	At4g40010; At5g66570; At1g11920; At2g45670; At1g14420; At1g05260; At2g41860; At1g05240
VII	sequence-specific DNA binding; regulation of transcription, DNA-dependent; regulation of RNA metabolic process; Transcription; transcription regulation; nucleus; transcription factor activity	At3g59600; At1g31880; At5g11530; At1g62886; At3g09250; At3g04570; At5g22500; At3g15510; At3g57670; At3g12910; At5g67411; At1g14410; At2g38300; At3g13890; At5g35770; At5g65640; At2g34820; At1g77200; At5g13330; At1g04370; At1g28300; At5g59340; At2g45050; At2g34830: At3g04100
VIII	transport; transmembrane region; integral to membrane	At5g66570; At1g25240; At1g07180; At5g49140; At3g01070; At2g45040; At2g41860; At1g52180; At3g54830; At5g59740; At1g04360; At3g13900; At2g38280; At1g52190; At1g30840; At1g01610; At1g01620; At2g38290; At3g60970; At1g80760; At2g39010; At1g31885; At3g04110; At5g46540
IX	floral organ development; post-embryonic organ development; flower development; embryonic development ending in seed dormancy; seed development; fruit development; reproductive structure development; reproductive developmental process; post-embryonic development	At3g57670; At1g31880; At5g35770; At3g63530; At2g41850; At3g13890; At5g59340; At2g38280; At2g34830; At1g28300
X	phenylpropanoid metabolic process cellular amino acid derivative metabolic process secondary metabolic process	At4g39650; At1g02940; At1g02930; At1g20510; At2g30210; At1g20500

XI	zinc ion binding; metal-binding; transition metal ion binding; metal ion binding; cation binding; ion binding	At2g41860; At2g45670; At1g11920; At3g13900; At1g14420; At3g01070; At1g04380; At1g02930; At5g66570; At3g09260; At3g57670; At3g27510; At1g01600; At1g05240; At1g05260; At2g30210; At2g45040; At1g04360; At2g38280; At3g60966; At2g45050; At2g19610; At3g27500; At5g22500; At3g63530; At1g25230
XII	response to water; response to water deprivation; response to abiotic stimulus	At5g66570; At5g22500; At3g04110; At3g09260; At4g40010; At1g05260; At1g01620; At1g02930
XIII	response to hormone stimulus; response to endogenous stimulus ; response to organic substance; intracellular signaling cascade; hormone-mediated signaling; activator; Ethylene signaling pathway; DNA-binding region:AP2/ERF; two-component signal transduction system (phosphorelay)	At1g04370; At1g77200; At5g13330; At1g31880; At3g13890; At4g27260; At3g27500; At5g22500; At1g62880; At2g45050
XIV	ATPase activity, movement of substances; hydrolase activity, acting on acid anhydrides; P-P-bond-hydrolysis-driven transmembrane transporter activity; primary active transmembrane transporter activity	At3g60970; At3g13900; At5g46540
XV	response to salt stress; response to osmotic stress; response to abiotic stimulus	At5g66570; At5g22500; At1g01620; At3g04110; At1g05260; At4g40010; At1g02930; At3g09260
XVI	Zinc finger, C3HC4 RING-type, zinc	At1g25230; At2g45040; At2g38280; At3g27500; At5g22500; At2g45050; At2g19610; At3g63530; At1g04360; At3g60966
XVII	heme binding; tetrapyrrole binding; electron carrier activity; iron ion binding; oxidation reduction	At3g01070; At1g25230; At2g30210; At5g22500; At1g75450; At1g04380; At1g05260; At1g05240; At1g01600
XVIII	Cyclin-like F-box; F-box associated type 1	At3g59590; At1g47810; At3g07870; At1g70970
XIX	macromolecule catabolic process; proteolysis; ubl conjugation pathway; modification-dependent protein catabolic process; proteolysis involved in cellular protein catabolic process; cellular macromolecule catabolic process	At2g45040; At2g43620; At4g39000; At2g43610; At4g39010; At4g38990; At4g11920; At1g04360; At1g75440
XX	response to light stimulus; response to radiation; response to abiotic stimulus	At1g05260; At4g40010; At1g01620; At3g09260; At1g02930; At5g66570; At5g22500; At3g04110
XXI	binding site:ATP; nucleotide phosphate-binding region:ATP; nucleotide-binding; purine ribonucleotide binding; ribonucleotide binding; adenyl nucleotide binding; purine nucleoside binding; nucleoside binding	At3g12915; At1g07180; At1g75450; At5g65690; At3g09240; At5g49140; At1g70980; At3g13900; At3g60970; At1g01460; At5g46540; At1g75440; At2g38280; At1g20500; At2g41860; At1g20510; At4g40010
XXII	response to fungus; disulfide bond; disulfide bond	At3g09260; At2g19893; At3g27503 At1g05240; At1g05260
XXIII	organelle lumen; membrane-enclosed lumen; intracellular organelle lumen	At3g59600; At3g63540; At5g66570
XXIV	chloroplast; transit peptide: Chloroplast	At5g66570; At3g63540; At5g11520
XXV	Protein kinase, core; protein amino acid phosphorylation phosphate metabolic process	At3g09240; At2g41860; At4g40010

3.1.4 Phenotypically outstanding mutants

The majority of the screened mutants phenotypically resembled wild-type plants when grown under green house conditions. Nevertheless, a few mutants exhibited an abnormal phenotype. Some mutants had mild phenotypes such as a slightly smaller size or darker leaves; these phenotypes were not always consistent and are therefore not described here. A strong phenotype on the other hand was observed for the mutant lines 57.31, 41.66, 72.29 and 94.12. These mutants will be described in this chapter.

3.1.4.1 Mutant line 57.31 exhibited an altered leaf phenotype

The mutant 57.31 had rounder shaped leaves than wild-type plants and a higher number of trichomes. In addition the plant was slightly darker than the wild-type plant (Fig. 22)

Fig. 22: Phenotype of mutant line 57.31 with increased number of trichomes. The plants were grown under green house conditions. (A) Col-*0* (B) mutant line 57.31.

The T-DNA insertion site of mutant line 57.31 was mapped on chromosome 5 at position 23932995. The insertion is between the genes At5g59330, which encodes for a putative lipid binding protein, and At5g59340, the transcription factor WUSCHEL-related homeobox 2 (WOX2) (Fig. 23). The mapping process identified further mutants with the same insertion site as mutant 57.31, these mutants are lines 70.24, 71.71, 73.30 and 70.06.

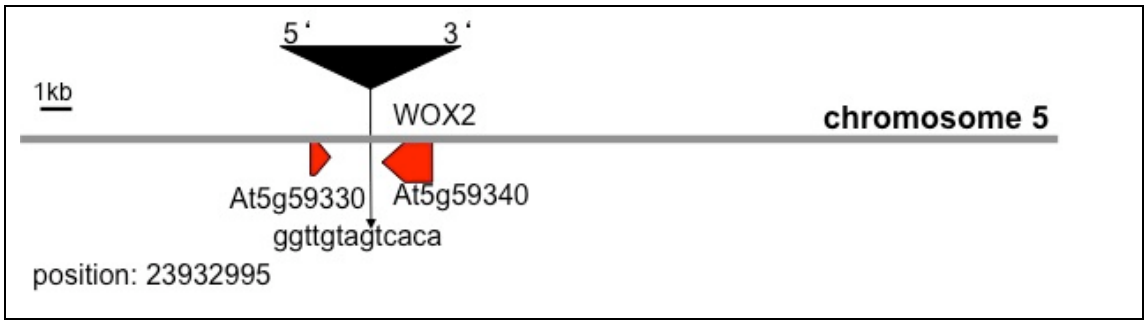


Fig. 23: Insertion site of mutant lines 70.24, 71.71, 57.31, 73.30 and 70.06.

Expression analyses of the genes flanking the insertion site revealed that the genes ultimately upstream and downstream of the insertion were up-regulated. The gene At5g59330 was up-regulated up to 9000 fold, while the gene At5g59330 was up-regulated up to 740 fold, in the tested cDNA samples. The phenotype could not be observed in all mutant lines. One explanation could be the varying expression of the influenced genes, generally observed within this population of activation tagged mutants. The results of subcellular localization studies with GFP of the protein encoded by gene At5g59330 are shown in the supplementary section.

3.1.4.2 Mutant line 41.66

Plants of the mutant line 41.66 were dwarfish and had curled leaves. This phenotype was also observed with the mutant lines 41.16, 41.58, 30.36 and 40.76. These lines have the same insertion site as mutant 41.66, although the phenotype was not as severe as in line 41.66. The mutant 41.66 will be described in detail in chapter 3.2.6.

3.1.4.3 Mutant line 72.29

The mutant 72.29 phenotypically differed from wild-type plants because the flowers of the mutant were missing the petals, while the sepals, stamen and carpel were developed (Fig. 24).

Fig. 24: Phenotype of the mutant line 72.29. The inflorescences of Col-*0* plants (A) and the mutant line 72.29 (B, C) are shown.

The line 72.29 was mapped because it led to an induction of the reporter systems. The mapping resulted in the identification of three insertion sites. The first insertion was localized in the third exon of gene At3g59590, a putative jacalin-type lectin. Expression analysis confirmed a down-regulation of this gene to 0.069 fold. The flanking genes At3g59580 and At3g59600 remained unaltered in their expression levels. This insertion site was confirmed via PCR analysis with primers flanking the insertion site. A segregation of the phenotype could be observed but this could not be ascribed to the genotype of the plants, when checking for homozygous and heterozygous individuals (data not shown). The mapping of line 72.29 was repeated and in addition to confirming the insertion site 1, two further insertion sites were identified. An insertion in the fourth exon of gene At5g59840, a Ras-related GTP-binding family protein, and in the third intron of gene At2g23140, an ubiquitin protein ligase, were found. Since several insertion sites were identified in this mutant line it is not possible to assign the lack of flower petals to one gene without further segregation analyses of all insertion sites. The segregation of the phenotype can either be ascribed to the genotype of the individual plants, if the loss of function of a gene is the cause, or, it can be led back to the variable expression levels of the flanking genes.

3.1.4.4 Mutant line 94.12

It was shown that mutant line 94.12 has the same insertion site as the mutant lines 94.52, 94.67 and 94.49. The insertion was localized in the first exon of gene At3g63840, which encodes for a protein of unknown function. The flanking gene At3g63530 encodes for an E3 ubiquitin ligase (also known as BIG BROTHER). This gene is a negative regulator of floral organ size. Studies have shown that a small

change in the expression level of BIG BROTHER can already alter organ size (*169*). The expression of gene At3g63530 was up-regulated 3.6 fold in the analyzed cDNA. The plants were dwarfish and showed abnormal leaf shape and development (Fig. 25).

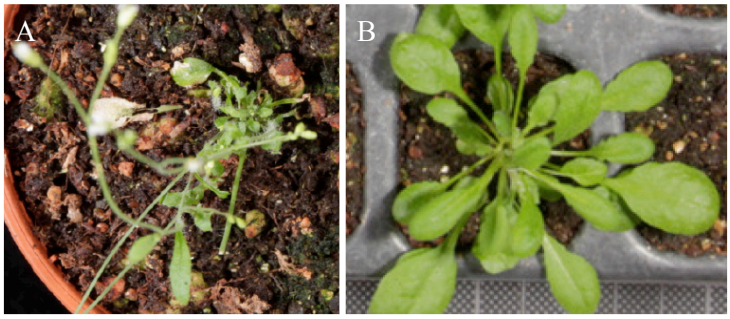


Fig. 25: Phenotype of mutant line 94.12. (A) Line 94.12 (B) Col-*0*. Grown under green house conditions. The picture was taken from C. Eigen, 2010 (*167*).

3.2 Analysis of Selected Genes

3.2.1 Selection and analysis of mutants

From the numerous genes discovered during the screening, candidates with putative roles in the production of chemoprotective metabolites were selected for further studies. This selection was based on the putative or known gene functions, the publicly available microarray data and the gene expression levels in the mutants. The putative function of the gene can give some insight into whether the gene is involved in plant secondary metabolism. Microarray data and coexpression networks of genes can be an indicator of a role in a certain metabolic pathway. Mutants that contain the T-DNA insertion in the coding region a of gene, leading to loss of function, and overexpression of upstream and downstream genes are more complicated to analyze in further studies than those with the T-DNA in the intergenic region resulting in the overexpression of only one or two genes. In order to assign the chemoprotective compound to a certain gene, both the knocked out gene and the overexpressed genes need to be subject of further analyses.

Subcellular localization studies were performed via fusion with GFP using the vector pGWB5 (C-terminal GFP fusion) or pGWB6 (N-terminal GFP fusion) and selected genes were cloned into pAMPAT for the generation of over expression lines (recapitulation lines) using the CaMV35S-promotor. The expression of GFP fusion proteins, on the one hand, can give some insight into the putative functions of still unknown proteins on the other hand it confirms the functionality of the clone. Fig. 26 summarizes the selection procedure of candidate mutants.

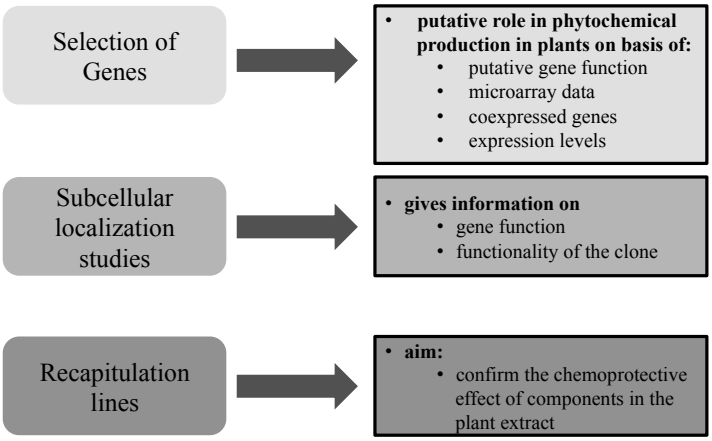


Fig. 26: Schematic diagram of the general procedure for selecting candidate mutants/genes and further analysis.

18 mutant lines were selected for further analyses, with the affected genes all having different predicted functions. The genes that were chosen for cloning, including the generated constructs and the respective mutants are listed in the following Table 12. The mutants 51.14, 52.18, 65.32, 41.66, 88.06 and 68.24 are dealt with in the following chapters. Additionally, GFP data is available for further genes not included in the results chapter, which can be found in the supplementary section. These are genes At4g37290, At4g37295, At1g01470, At5g59330, At5g49120 and At1g08670 (see chapter 5). The following table lists all constructs developed in course of the analysis of chemoprotective TAMARA mutants. The constructs marked with "*" were created and analyzed by L. Marbaise. These data are published in her Master's Thesis (2010) (*170*).

Table 12: List of constructs

Mutant	Gene	Entry vector	Expression vector
52.18	At3g54830	clone 5 (pDONR207) clone 6 (pDONR207) clone 2.9 (pDONR207) V55: clone2.9 (in frame) (pDONR207) V15: promshortAt3g54830 (pDONR207)	V16: clone5 (pGWB6) V21: clone5 (pAMPAT) V17: clone6 (pGWB6) V22: clone6 (pAMPAT) V18: clone2.9 (pGWB6) V23: clone2.9 (pAMPAT) V58: clone2.9 (pGWB5) V20: promshortAt3g54830 (pGWB3i)
65.32	At1g63820	V11: At1g63820 (pDONR207)	V13: At1g63820 (pGWB5) V25: At1g63820 (pAMPAT)
47.64	At4g37290 At4g37295	V40: At4g37290 (pENTR1A) V41: At4g37295 (pENTR1A)	V49: At4g37290 (pGWB6) V62: At4g37290 (pAMPAT) V50: At4g37295 (pGWB5) V63: At4g37295 (pAMPAT)
88.6	At5g55880	V33: At5g55880 (pENTR1A) 7 *: N-terminal At5g55880 (pDONR207)	V44: At5g55880 (pGWB5) V59: At5g55880 (pAMPAT) 7 *: N-term.At5g55880 (pGWB5)
51.14	At3g09260	V34: At3g09260 (pENTR1A)	V45: At3g09260 (pGWB5) V60: At3g09260 (pAMPAT) V79: At3g09260 (pGWB8)
63.16	At1g08670	V43: At1g08670 (pENTR1A)	V51: At1g08670 (pGWB5) V53: At1g08670 (pDEST24) V54: At1g08670 (pDEST42) V61: At1g08670 (pAMPAT)
51.70	At1g01470	V35: At1g01470 (pENTR1A)	V46: At1g01470 (pGWB5) V64: At1g01470 (pAMPAT)
70.06	At5g59330	V36: At5g59330 (pENTR4)	V47: At5g59330 (pGWB5) V65: At5g59330 (pAMPAT)
70.70	At5g49120	V37: At5g49120 (pENTR4)	V52: At5g49120 (pGWB5) V66: At5g49120 (pAMPAT)
68.24	At1g71002	43*: At1g71002 (pDONR207) V72: promshortAt1g71002 (pENTR1A) V76: prommidAt1g71002 (pENTR1A)	V78: promshortAt1g71002 (pGWB3i) V80: prommidAt1g71002 (pGWB3i) 43*: At1g71002 (pGWB2) 43 (V): At1g71002 (pAMPAT)
90.46	At5g09520 At5g09540	25a*: At5g09520 (pDONR207) 25b*: At5g09540 (pDONR207)	25a*: At5g09520 (pGWB5) 25a (V): At5g09520 (pAMPAT) 25b*: At5g09540 (pGWB5) 25b*: At5g09549 (pAMPAT)
87.58	At5g10190	32*: At5g10190 (pDONR207)	32*: At5g10190 (pGWB5) 32*: At5g10190 (pAMPAT)
81.77	At2g37440	52*: At2g37440 (pDONR207)	
64.22	At2g19893	63*: At2g19893 (pDONR207)	63*: At2g19893 (pGWB5) 63*: At2g19893 (pAMPAT)
64.48	At1g70970	65*: At1g70970 (pDONR207)	65*: At1g70970 (pGWB5) 65*: At1g70970 (pAMPAT)
72.14	At1g02930	67*: At1g02930 (pDONR207)	67*: At1g02930 (pGWB5) 67*: At1g02930 (pAMPAT)
59.67	At1g05260	69*: At1g05260 (pDONR207)	69*: At1g05260 (pGWB5) 69 (V): At1g05260 (pAMPAT)
50.46	At2g43610	72a*: At2g43610 (pDONR207)	72a *: At2g43610 (pGWB5) 72a (V): At2g43610 (pAMPAT)

Clones marked with "*" were constructed by Lillian Marbaise and published in her Master's Thesis in 2010 (*170*).

3.2.2 Mutant lines 51.14, 42.25 and 66.71

3.2.2.1 The insertion site of mutant lines 51.14, 42.25 and 66.71 was mapped on chromosome 3

In three individual mutant lines (51.14, 42.25 and 66.71) the insertion was localized on chromosome 3 in the third exon of a gene encoding for a DNA-binding protein/nuclease. The neighboring genes encode for the β- glucosidase PYK10 (At3g09260), and a putative protein kinase (At3g09240) (Fig. 27). The mutant was named *PYK10-1D-cp* (PYK10-dominant-chemoprotective). In expression analyses of RNA, derived from leaf material of line 51.14, the gene encoding for PYK10 was up-regulated up to 5,500 fold. The gene encoding for the protein kinase was up-regulated 17 fold.

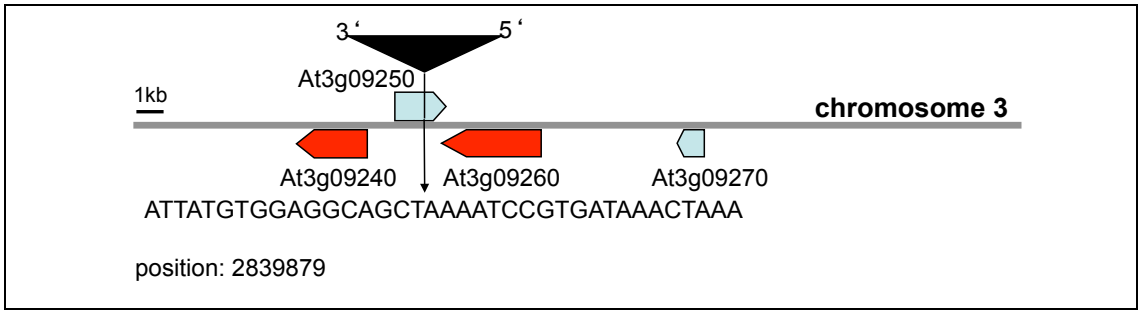


Fig. 27: Insertion site of mutant lines 51.14, 42.25 and 66.71.

PYK10 encodes for a root and hypocotyl specific β-glucosidase and belongs to the family of glycosyl hydrolases 1 (*171,172*). It was demonstrated that PYK10 exhibits β-D-glucosidase and β-D-fucosidase activity (*173*). So far it was shown that PYK10 hydrolyzes the hydroxycoumarin scopolin and esculin *in vitro* (*174*). The coexpression network from the Atted-II database of PYK10 is shown in Fig. 28 (*175, 176*). The monooxygenase CYP81F4 is coexpressed with PYK10 (Fig. 28). This enzyme has been shown to synthesize 1MO-I3M (*177*). PYK10 is coexpressed with the transcription factor MYB34, which regulates the production of 1MO-I3M in roots (*178*).

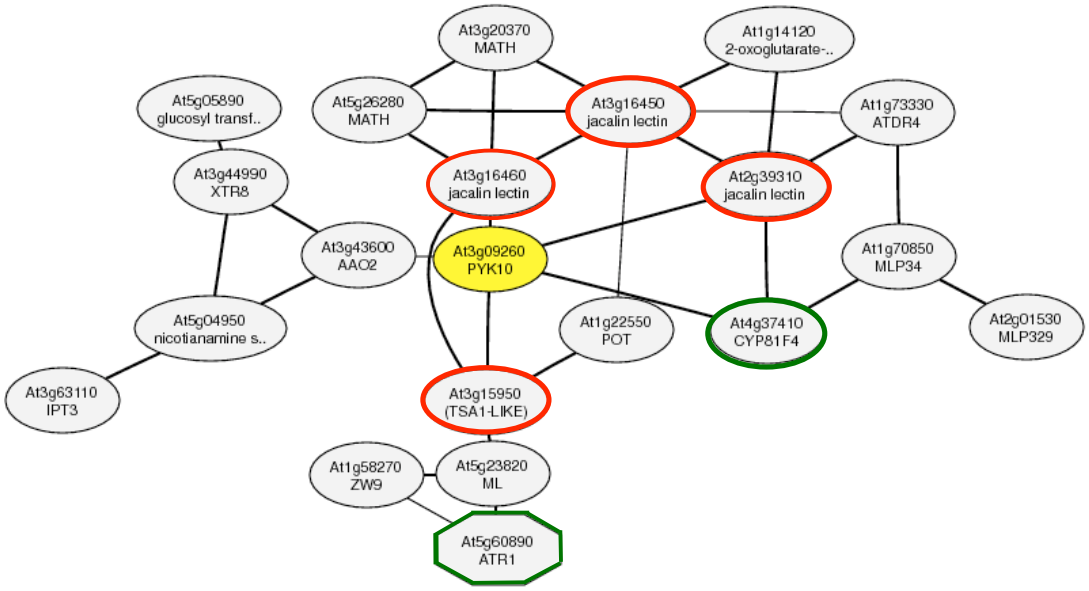


Fig. 28: Coexpression network of gene At3g09260 (Atted-II database). Jacalin lectin= homologs of nitrile specifier proteins, ATR1=MYB34, TSA1LIKE= NAI2.

3.2.2.2 Cloning and subcellular localization of PYK10

The PYK10 gene was cloned into the entry vector pENTR1A and the gateway cassette containing the gene was recombined into the expression vector pGWB5, for the production of C-terminal GFP fusion protein, and PAMPAT, for the over expression of PYK10 *in planta*. PYK10 protein is known to be the main component of constitutive endoplasmatic reticulum (ER) bodies, which are spindle shaped and surrounded by ribosomes (*173,179*). This could be confirmed via expression of the construct 35S:PYK10-GFP in *A. thaliana* cultured cells (Fig. 29). The ER bodies were visible as spindle shaped structures throughout the cell lumen. The expression and correct localization of the GFP-tagged PYK10 protein was used as a control to confirm the functionality of the PYK10 clone.

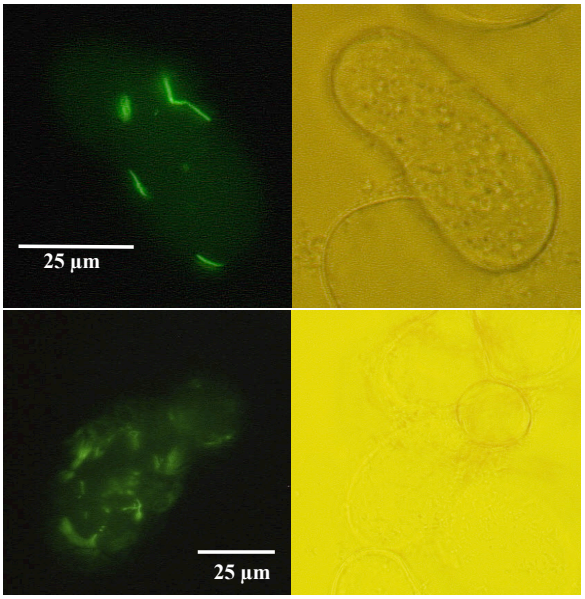


Fig. 29: Wide-field fluorescence microscopy of *A. thaliana* cultured cells transformed with 35S:PYK10-GFP.

3.2.2.3 Isolation of homozygous PYK10 insertion lines

The T-DNA insertion line WiscDsLox461-464F2 was obtained from NASC. This line contains the T-DNA insertion in the first intron of At3g09260 (Fig. 30). Nagano *et al*. (2008) have previously shown by immunoblot analysis that the line WiscDSLOX461-464F2 (*pyk10-1*) has no PYK10 protein (*180*). Homozygous plants of line WiscDsLox461-464F2 were isolated (Fig. 31).

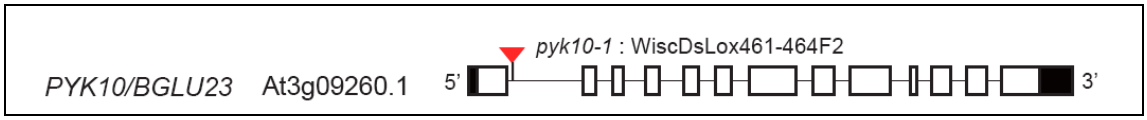


Fig. 30: Insertion site of T-DNA insertion line WiscDsLox461-464F2. The figure was taken from the publication Nagano *et al.* (2008) (*180*).

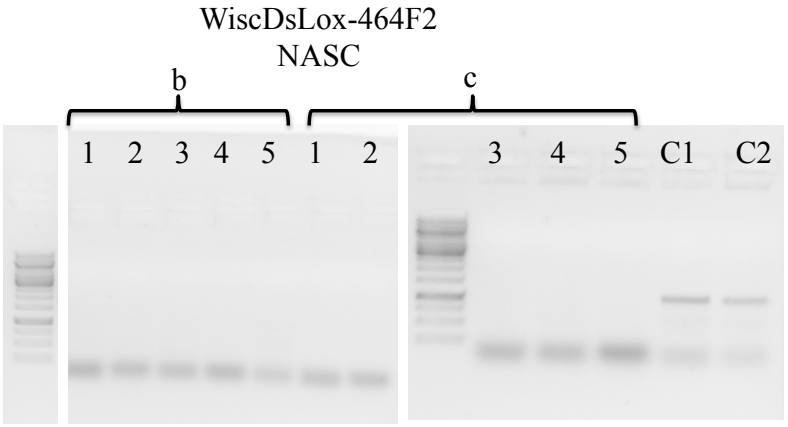


Fig. 31: Agarose gelelectrophoresis of a PCR reaction using genomic DNA of PYK10 knockout lines as template DNA and primers flanking the T-DNA insertion site. Homozygous knockout lines of PYK10 (WiscDsLox461-464F2) were identified. The plants were obtained from NASC. The T-DNA insertion site is located in the first intron at position: 2843529. The expected PCR-fragment has a size of 1080 nt. C1, C2= control plants (Col-*0*).

3.2.2.4 Production of 35S:PYK10 lines

The 35S:PYK10 construct in pAMPAT vector was used for stable transformation of Col-*0* plants. 26 lines were analyzed for their expression in the T_1 generation. Among these 26 lines, 14 overexpressing lines could be identified. The lines that showed the highest expression were chosen for further analysis, namely lines 59, 43 and 19.

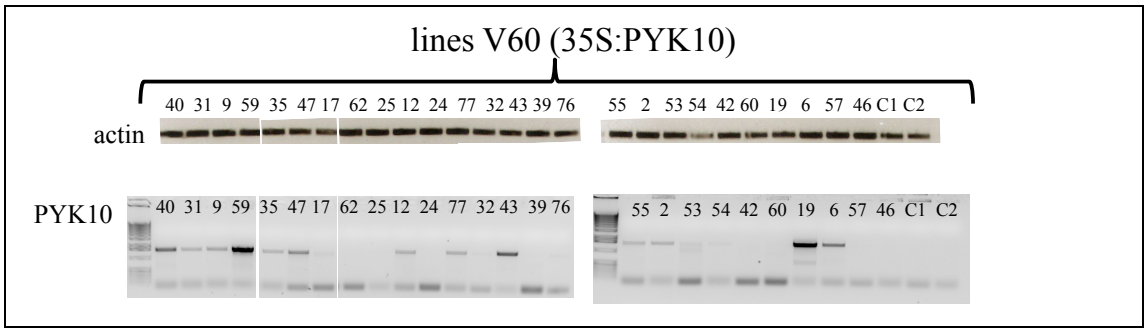


Fig. 32: Agarose gelelectrophoresis of the results of a semi quantitative PCR analysis of PYK10 expression in T_1 plants transformed with 35S:PYK10. Upper panel shows the actin control PCR. The lower panel the expression of PYK10 using gene specific primers. The actin PCR was performed with 25 cycles, the PYK10 PCR with 40 cycles. The full-length coding region was amplified, which has a length of 1575 nt.

3.2.2.5 Confirmation of chemoprotective activity in PYK10 overexpressing lines using Hepa1c1c7-Lux reporter cells

The three PYK10 overexpressing lines 59, 43 and 19 were chosen for confirmation of the gene-to-trait relationship. Two samples were taken for each line. Each sample consisting of leaf material pooled from two plants. Methanol extracts were made from

these pools. The DMSO samples were diluted 1:5 in cell culture medium. Two µL of diluted plant extracts were used to induce the cell culture. The two sample pools taken for each line were tested in two individual screens (screen 1 and screen 2). The third screen was performed with the samples from screen 1, which were diluted once more to 1:50, in order to exclude any inhibiting effect of a high amount of active compounds in one sample (see 3.1.1). During each screening assay three Col-*0* wt samples were used as control. The median of all triplicates of control measurements was determined and used as a standard. The activation tagged line 51.14 and *pyk10-1* were tested in parallel. In a first screen the lines 51.14, V60.59 and *pyk10-1* led to an increased luminescence signal in comparison to Col-*0* extracts. These results were confirmed in screen 3. Screen 2 confirmed the inducing effect of the extracts from 51.14, V60.59 and *pyk10-1*, here also the extracts from lines V60.19 and V60.43 led to a higher luminescence signal in comparison to wild-type plants. All in all the extracts from plants overexpressing PYK10 had an inducing effect on the reporter cells. The results suggest that the enhanced expression of PYK10 in the activation tagged lines leads to the production of EPRE inducing metabolic compounds. In addition, the loss of function of PYK10 led to an induction of EPRE. Mutant *pyk10-1* probably accumulates compounds that are normally converted by PYK10.

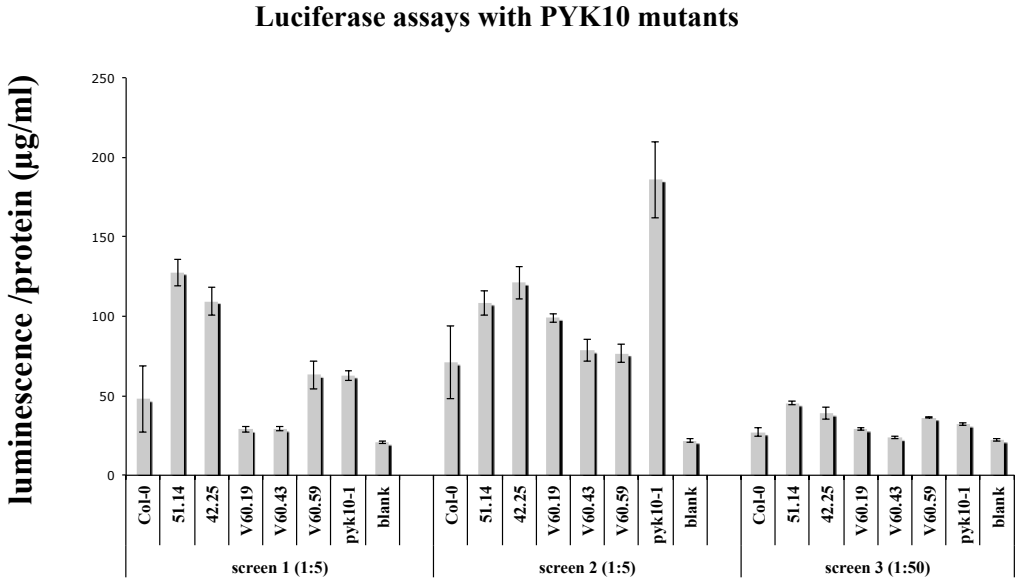


Fig. 33: Analysis of the induction potential of PYK10 mutants in Hepa1c1c7-Lux cells. Under analysis were the recapitulation lines (V60), line 51.14 and *pyk10-1*. The luminescence signal is shown as the median luminescence per total cellular protein (µg/mL) of three technical replicates per biological replicate (screen 1 and 2). Screen 3 consists of 1:50 diluted samples of screen 1. The bars represent the standard error of 3 technical replicates for the mutants. The Col-*0* median was determined from three to four samples per screening assay with three technical replicates each.

3.2.2.6 Metabolite analysis of root extracts of PYK10 mutants

The secondary metabolites of the PYK10 mutants *PYK10-1D-cp*, *35S:PYK10* and *pyk10-1* were analyzed by Dr. Christoph Böttcher (Leibniz Institute of Plant Biochemistry, Halle). The samples were taken from ca. 4-week-old seedlings, which were grown vertically on ½ MS-plates with gelrite, under long day conditions in the growth cabinet. Since PYK10 is mainly expressed in roots, extracts of roots were analyzed. Each sample consisted of three pooled seedlings from one plate. The root extracts were analyzed for 1MO-I3M and 4MO-I3M and their respective nitrile (1-methoxyindol-3-acetonitrile (1MO-IAN), 4-methoxyindol-3-acetonitrile (4MO-IAN)) and ascorbigen breakdown products (1-methoxyindol-3-ascorbigen (1MO-ascorbigen), 4-methoxyindol-3-ascorbigen (4MO-ascorbigen)). PYK10 is hypothesized to have myrosinase activity. Since it was demonstrated *in vitro* that PYK10 hydrolyzes scopolin and esculin, the content of scopolin was also determined. Esculin could not be detected in the given samples. The 1MO-I3M content was 30 to 50% reduced in the analyzed *PYK10-1D-cp* and the *35S:PYK10* lines in comparison

to Col-*0*. In the *pyk10-1* line the 1MO-I3M content was comparable to the content in Col-*0* (Fig. 34 A). The 1MO-ascorbigen content was reduced in the *PYK10-1D-cp* and the *pyk10-1* lines and unaltered and in the *35S:PYK10* lines in comparison to Col-*0* (Fig. 34 B). The 1MO-IAN content was strongly decreased in the *pyk10-1* line to around 30% of the content in Col-*0*. In the *PYK10-1D-cp* and the *35S:PYK10* lines it was comparable to the 1MO-IAN content in Col-*0* (Fig. 34 C).

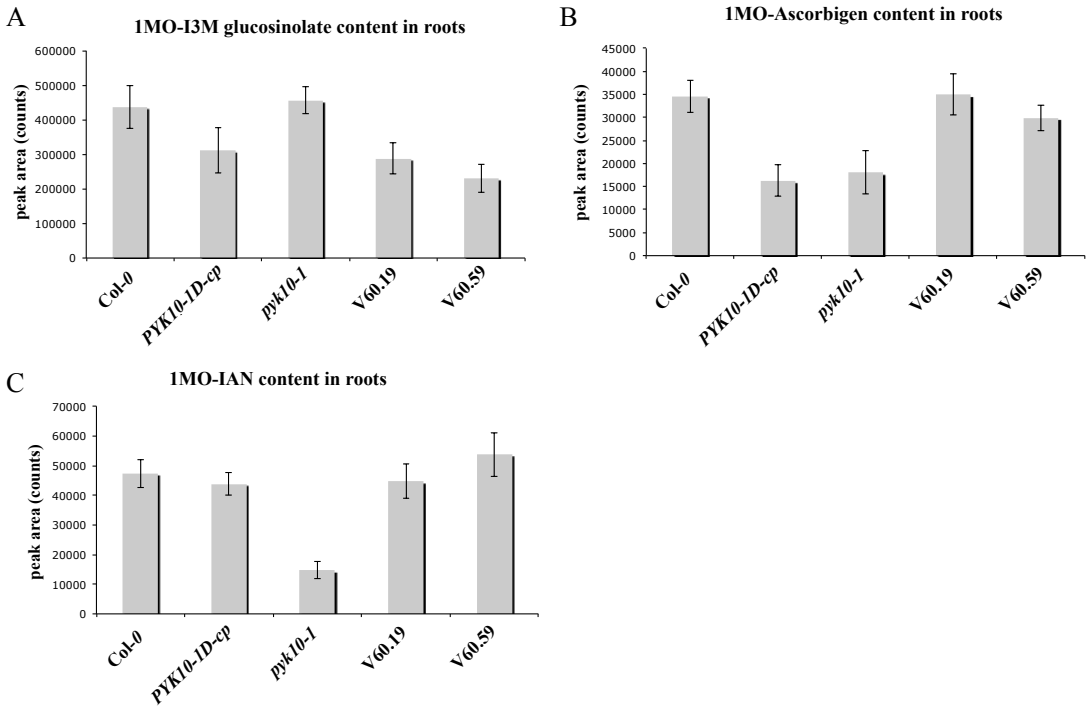


Fig. 34: UPLC-ESI-QTOF-MS analysis of 1MO-I3M and its derived compounds in PYK10 mutants. (A) 1-methoxyindol-3-ylmethylglucosinolate (1MO-I3M) and the breakdown products (B) 1-methoxyindol-3-ascorbigen (1MO-ascorbigen) and (C) 1-methoxyindol-3-acetonitrile (1MO-IAN) in roots of four-week-old *A. thaliana* mutant seedlings. Under analysis were the mutants *pyk10-1*, *pyk10-1D-cp*, and the *35S:PYK10*-lines V60.19 and V60.59. The median of 10 technical replicates of each mutant (6 for V60.19) was determined along with the standard error (bars).

The 4MO-I3M content in roots was comparable to Col-*0* in all mutants. It was mildly decreased in the overexpressing lines *35S:PYK10* (Fig. 35 A). The 4MO-ascorbigen content was slightly decreased in the *pyk10-1* and *PYK10-1D-cp* lines (Fig. 35 B). In the *pyk10-1* mutant the 4MO-IAN content was around 16% of then content found in Col-*0*. In the *PYK10-1D-cp* the nitrile content was 1.2 fold increased and in the *35S:PYK10* V60.19 line it was decreased to 74% of Col-*0* (Fig. 35 C). The scopolin content is shown in Fig. 35 D. Here the content was mildly reduced in all analyzed mutants in comparison to Col-*0*.

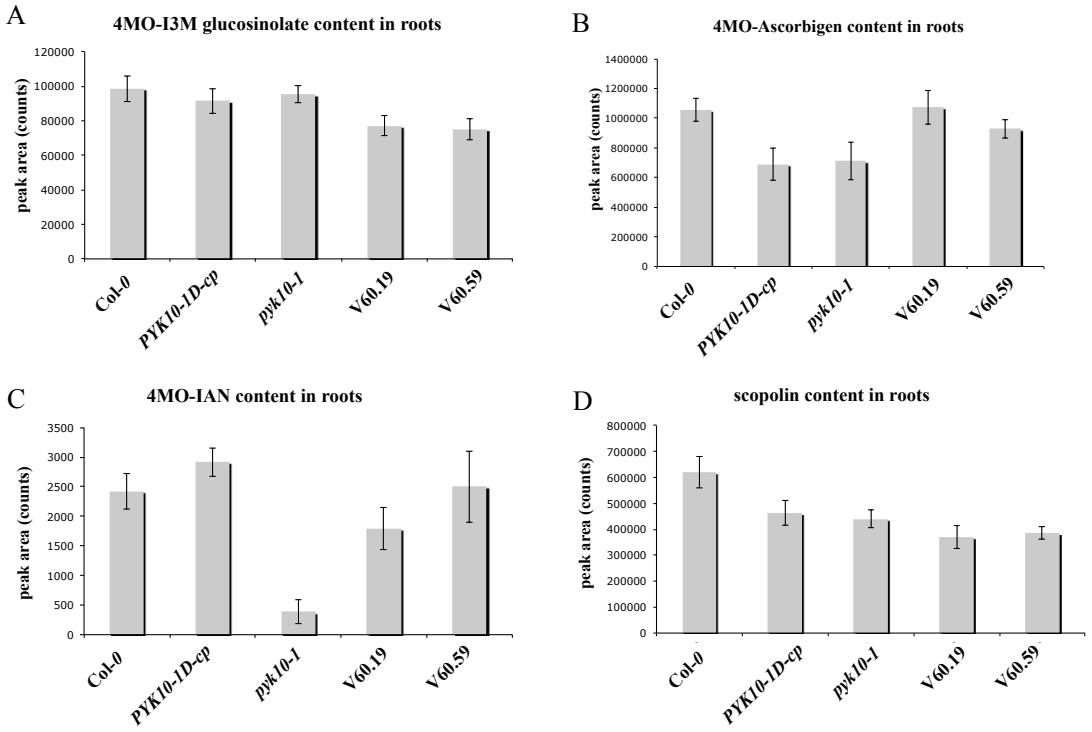


Fig. 35: UPLC-ESI-QTOF-MS analysis of 4MO-I3M and its derived compounds in PYK10 mutants. (A) 4-methoxyindol-3-ylmethylglucosinolate (4MO-I3M) and the breakdown products (B) 4-methoxyindol-3-ascorbigen (4MO-ascorbigen) and (C) 4-methoxyindol-3-acetonitrile (4MO-IAN) in roots of four-week-old *A. thaliana* mutant seedlings. Under analysis were the mutants *pyk10-1*, *pyk10-1D-cp*, and the *35S:PYK10*-lines V60.19 and V60.59. The median of 10 technical replicates of each mutant (6 for V60.19) was determined along with the standard error (bars).

In summary the data suggest that PYK10 overexpression leads to mildly reduced amounts of 1MO-I3M in the roots. The loss of function mutant produces less 1MO-IAN and 4MO-IAN.

3.2.3 Mutant lines 52.18, 46.23 and 46.24

3.2.3.1 The insertion site of mutant lines 52.18, 46.24 and 46.23 was mapped on chromosome 3

The T-DNA insertion of mutants 52.18, 46.23 and 46.24 was mapped on chromosome 3 between the genes At3g54830, a putative amino acid transporter, and At3g54840, a RAB-like GTPase (Fig. 36). The mutant was named *AAT-1D-cp* (Amino acid transporter-dominant-chemoprotective). An up-regulation of up to 80-

fold of the expression in comparison to wild-type plants was determined for gene At3g54830 in the analyzed cDNAs, whilst the expression of At3g54840 was unaltered.

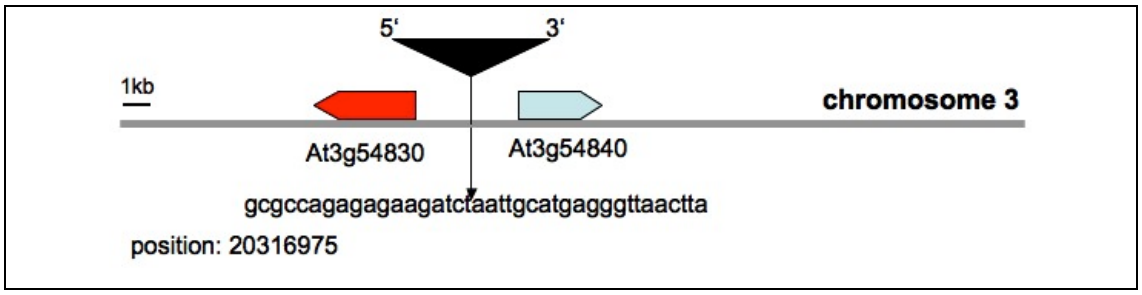


Fig. 36: Insertion site of mutant lines 52.18, 46.23 and 46.24.

At3g54830 is annotated in TAIR as a putative amino acid transporter based on sequence and structure similarities (*158*). *In silico* analysis of At3g54830 using the tool ARAMEMNON (*159*) shows that this putative amino acid transporter is clustered together with ten other putative amino acid transporters. On the basis of sequence homology, At3g54830 has been grouped into the amino acid/ auxin permease family. Furthermore, two related amino acid transporters (At5g65990 and At5g02180) have been subcellularly localized to the endomembrane system (*181*). The prediction tool THMHMM was applied in order to predict the topology of the putative transporter (*182*). The protein of 546 amino acids (AA) in length and 59.7 kDa molecular mass has six putative transmembrane domains.

3.2.3.2 *AAT-1D-cp* accumulated isothiocyanates

Metabolic profiling using UPLC-ESI-QTOF-MS performed and analyzed by Dr. Stephan Schmidt (Leibniz Institute of Plant Biochemistry, Halle) indicate that *AAT-1D-cp* accumulates the glucosinolate I3M, ascorbigen, which is a breakdown product of glucobrassicin that is formed in the presence of ascorbic acid, the flavonoid kaempferol, and the ITCs 4-methylsulphinylbutyl (4-MSOB-ITC), 5-methylsulphinylpentyl (5-MSOP-ITC), 7-methylsulphinylheptyl (7-MSOH-ITC) and 8-methylsulphinyloctyl (8-MSOO-ITC). The chromatogram excerpt in Fig. 37 illustrates the detected ITCs.

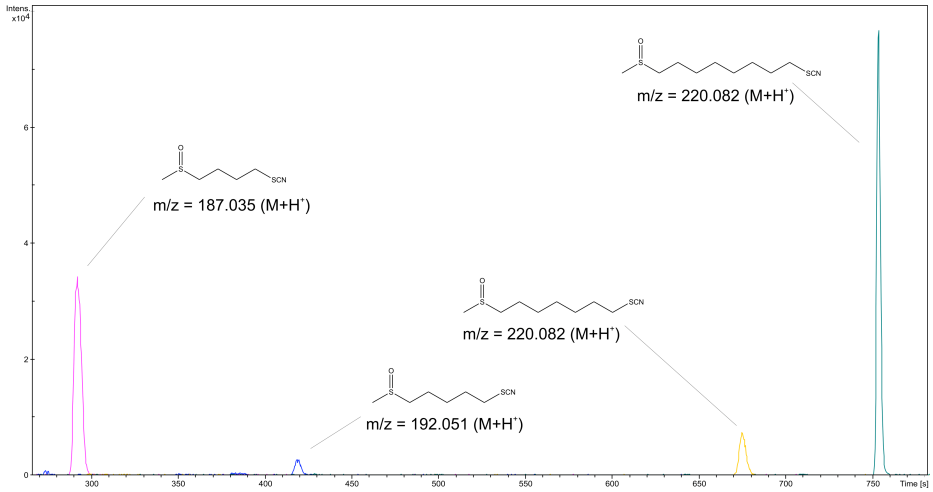


Fig. 37: Excerpt of the chromatogram of mutant *AAT-1D-cp*. The mutant accumulates different ITCs. The figure was provided by Dr. Stephan Schmidt (IPB, Halle).

3.2.3.3 Alternative splicing of gene At3g54830

The gene At3g54830 was amplified from cDNA, derived from both mutant and Col-*0* RNA, and cloned into the gateway vector pDONR207. Sequencing analyses indicate that the mRNA is spliced in alternative ways. On the basis of the sequencing results it can be concluded that the mRNA of At3g54830 exists in at least three different splicing variants. These were named after the clones 5, 6 and 2.9. Based on ARAMEMNON, the gene is predicted to have 12 exons (*159*). The first seven exons of the three different splicing variants correlate with the annotated sequence. The alternative splicing of the mRNA, if translated, would lead to three different proteins. From the sequencing results it can be predicted that clone 5 results in a protein with 543 AA, clone 6 with 448 AA and clone 2.9 with 451 AA. Both for clone 5 and clone 2.9 the new stop codon is 16 nt downstream of the original stop codon. The clone 6 has a predicted premature stop codon. Nine cDNAs derived from Col-*0* RNA of leaves were sequenced and from these nine clones, six correlated with clone 5, clone 6 was detected twice and the splicing variant of clone 2.9 was detected once. A schematic diagram of the splicing variants is shown in Fig. 38.

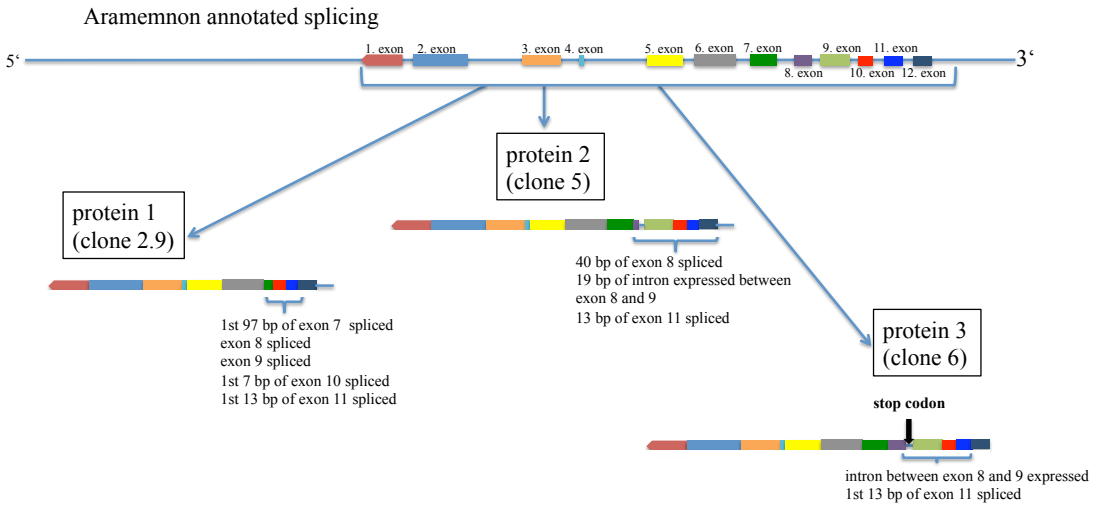


Fig. 38: Schematic diagram of the detected splicing variants of At3g54830 mRNA.

3.2.3.4 The putative amino acid transporter was localized to the endomembrane system

For the visualization of the subcellular localization of the three different cDNA clones, N-terminal GFP-fusion proteins were constructed using the expression vector pGWB6. The alternate splicing of the putative transporter C-terminal fusion with GFP led to a frame shift of the reporter gene. Therefore, the GFP was fused N-terminally to the different splicing variants. The fusion proteins were transiently expressed in *A. thaliana* cultured cells. GFP fluorescence was detected in cells transformed with clone 5 (Fig. 40) and clone 2.9 (Fig. 39) but not with clone 6 (data not shown). The fluorescence signal indicated a localization of the putative transporter in the endomembrane system, presumably in the ER.

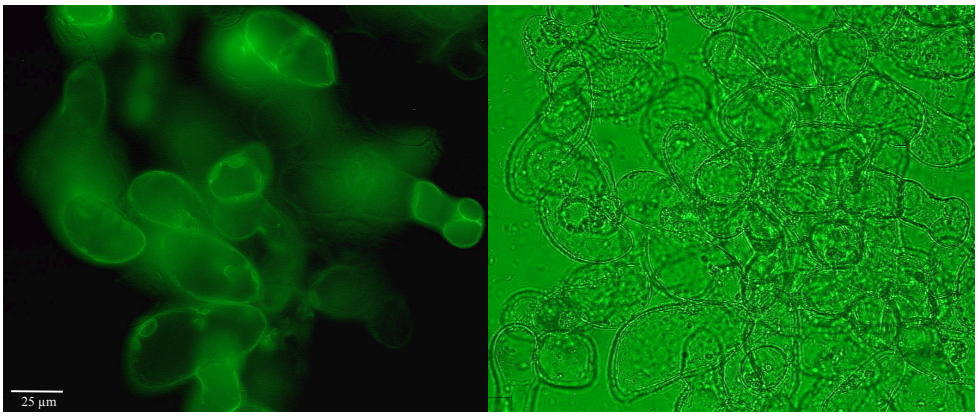


Fig. 39: Wide-field fluorescence microscopy of *A. thaliana* cultured cells transformed with 35S:GFP-At3g54830 (clone 2.9).

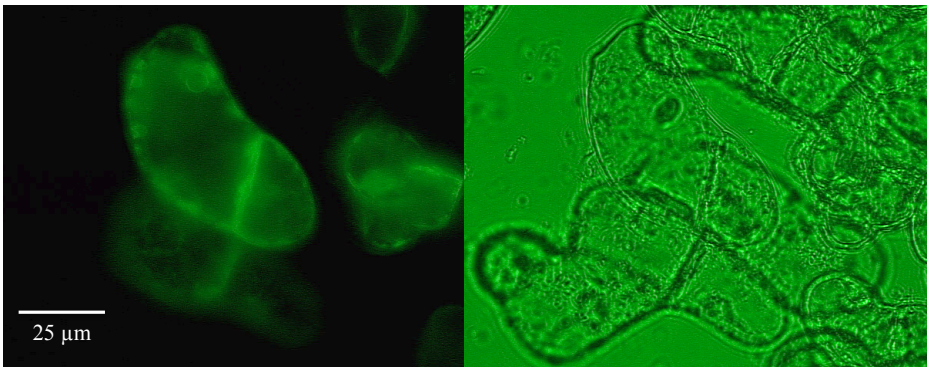


Fig. 40: Wide-field fluorescence microscopy of *A. thaliana* cultured cells transformed with 35S:GFP-At3g54830 (clone 5).

To confirm an ER-localization of the putative transporter, the transformed *A. thaliana* cultured cells were treated with BFA (2.7.2.1). This toxin disrupts the Golgi apparatus and leads to mislocalization of ER resident protein as well. After 2 h of treatment with the cytotoxin, the distinct fluorescence structures that could be observed before the treatment had disappeared. The GFP signal could be seen throughout the cell lumen. This observation was made for both clones (Fig. 41)

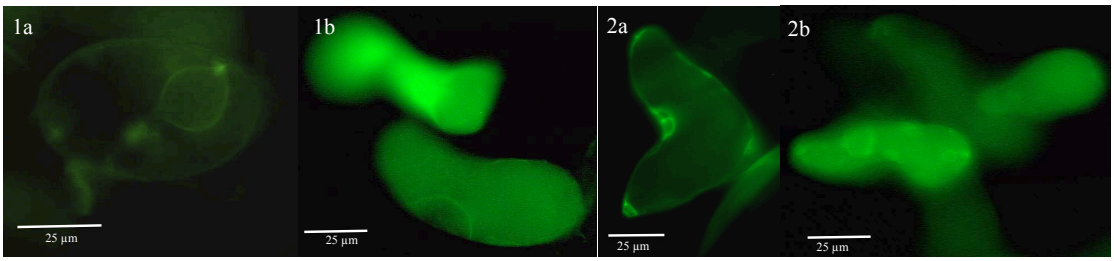


Fig. 41: Wide-field fluorescence microscopy of *A. thaliana* cultured cells treated with BFA and transformed with 35S:GFP-At3g54830. *A. thaliana* cultured cells transiently expressing clone 2.9 (1) and clone 5 (2). (a) untreated cells (b) with BFA (100 µg/mL; 2 h; 22 °C) treated cells.

3.2.3.5 Promotor analysis

To find out more about the role of At3g54830 in the plant, 1992 nt of the intergenic region and the first three exons of At3g54830 were cloned via gateway cloning into pDONR207. The GUS gene was placed under the control of the putative promotor region using the expression vector pGWB3i. GUS expression was analyzed in seedlings and adult plants of the T_2 generation of the stably transformed *A. thaliana* plants of three independent lines. The GUS signal was observed in the veins of adult leaves (green house conditions). GUS activity was also visible in the cotyledons, the leaves and in root tips of seedlings.

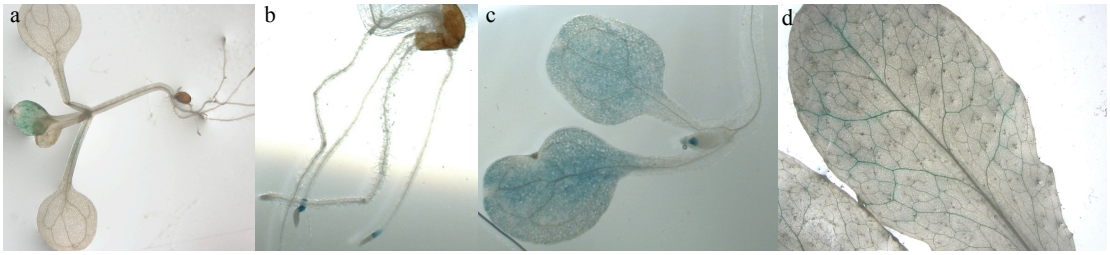


Fig. 42: GUS analysis of promAAT:GUS in *A. thaliana* Col-*0* plants. (a) 13-day-old seedling (b) 10-day-old seedling roots (c) 10-day-old seedling cotyledons, (d) adult leaf (T_2). The plants from (a), were grown on ½ MS-plates under long day conditions, (b) and (c) under short day conditions. The adult plants were grown in the greenhouse (d).

3.2.3.6 Production of At3g54830 overexpressing lines and analysis in Hepa1c1c7-Lux cells

To obtain further information on the functional splicing variant of At3g54830, the different clones were placed under the control of the CaMV35S-promotor in the pAMPAT vector. Different lines of selected T_1-generation plants were tested in the Hepa1c1c7-Lux cells. In general an induction of the luminescence could not be observed (data not shown).

3.2.4 Mutant lines 65.32 and 67.18

3.2.4.1 The insertion site of mutant lines 65.32 and 67.18 was mapped on chromosome 1

The T-DNA insertion site of mutants 65.32 and 67.18 was mapped on chromosome 1 between the genes At1g63810 and At1g63820 (Fig. 43). An up-regulation of At1g63820 of up to 53-fold to the expression in wild-type plants was determined. The expression of the gene At1g63810 was not changed. The gene At1g63810 encodes for a putative Nrap-like nucleolar RNA-associated protein. The function of the protein encoded by At1g63820 is unknown. *In silico* analysis using ARAMEMNON predicts that this unknown protein is soluble. It contains a CCT-domain (CONSTANS, CONSTANS-like, TOC1), which is a short motif of 43 amino acids containing a putative nuclear localization signal within its second half. The protein is 51% homologous to one other protein (At5g41380), which is of unknown function as well. Maximum expression of At1g63820 is in dry and germinating seeds (*183*) (Fig. 44). The protein belongs to the ASML2 (Activator of Spo^{min}::LUC2)

family proteins class I. One member of this protein family, ASML2 (At3g12890), has been shown to function as a transcriptional activator. ASML2 regulates a subset of sugar inducible genes (*184*).

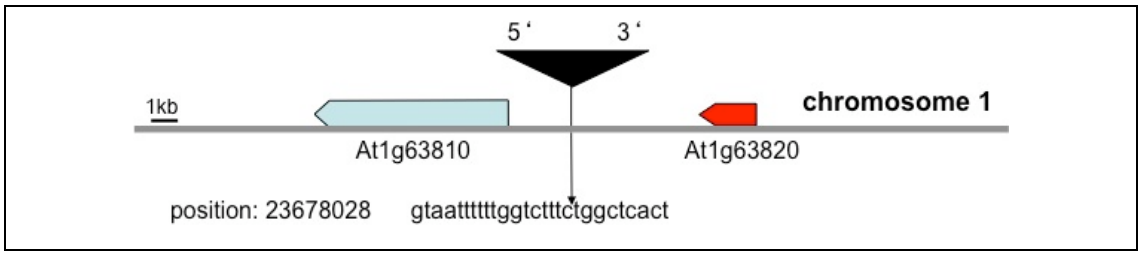


Fig. 43: Insertion site in mutant lines 65.32 and 67.18.

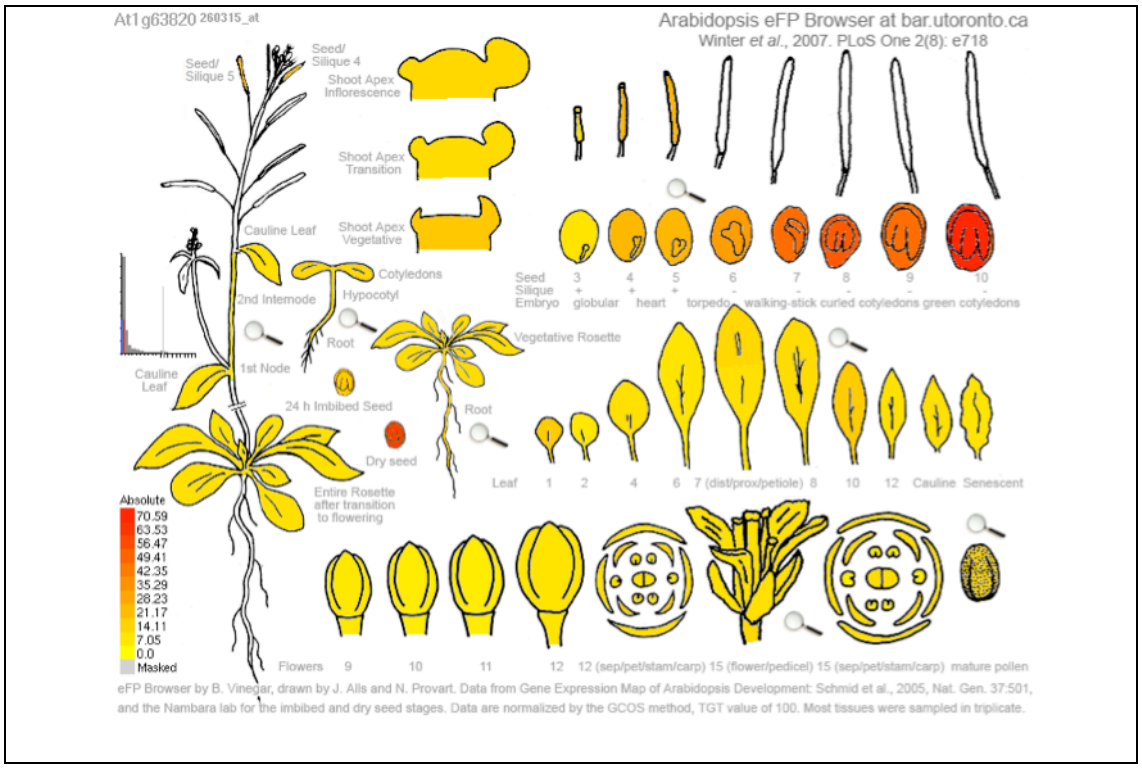


Fig. 44: Expression level of At1g63820 in different Arabidopsis organs. The gene is expressed in the siliques and seeds, but only mild or no expression takes place in the leaves and other organs. The highest expression takes place in seeds (stages 7 to 10). Source: Arabidopsis eFP Browser (*183*).

3.2.4.2 Subcellular localization analysis of gene At1g63820

To determine the subcellular location of the protein it was C-terminally fused with GFP. Both in *N. benthamiana* cells and in *A. thaliana* cultured cells the transient expression led to a nuclear and cytoplasmic GFP signal (Fig. 45). These results confirmed the prediction for the CCT domain.

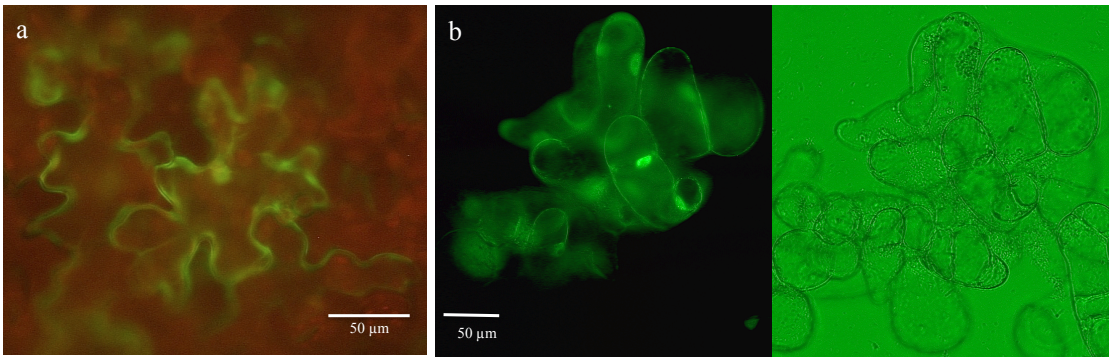


Fig. 45: Wide-field fluorescence microscopy of GFP fluorescence of cells transiently expressing 35S:At1g63820-GFP. (a) *N. benthamiana* leaves (b) *A. thaliana* cultured cells.

3.2.4.3 Production of At1g63820 overexpressing lines and analysis in Hepa1c1c7-Lux cells

The gene At1g63820 was stably expressed in *A. thaliana* plants under the control of the CaMV35S-promotor using the pAMPAT vector. Three overexpressing lines (V25.5, 7 and 9) were tested in the Hepa1c1c7-Lux reporter system.

Fig. 46 is a summary of the screening results in Hepa1c1c7-Lux cells. The line 65.32 was used for comparison. In both screening assays the lines 25.7 and 25.9 showed a high induction of the reporter system. The TAMARA line 65.32 did not induce the system as strongly as in past screening with an induction of 1.27 and 1.95 fold to wild-type plants. Since the expression levels of the genes in the activation tagged lines was not stable, this was observed in many TAMARA lines (data not shown), the accumulation of active compounds in these mutants may also vary leading to variable induction potential in the reporter systems.

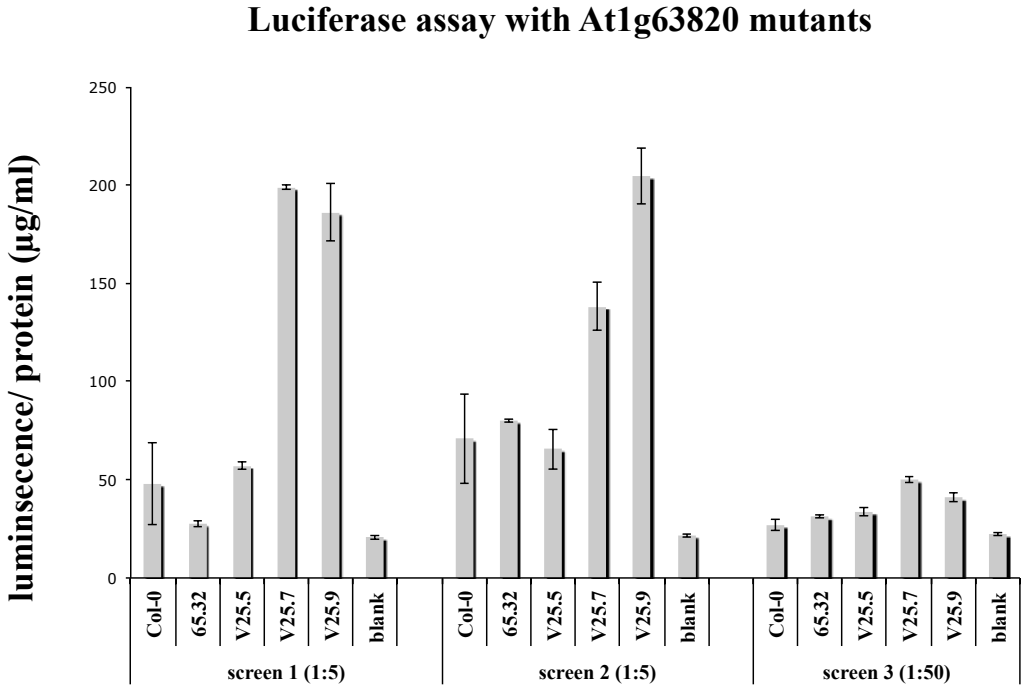


Fig. 46: Analysis of the induction potential of At1g63820 recapitulation lines (V25) using the Hepa1c1c7-Lux reporter system. The luminescence signal is shown as median luminescence per total cellular protein (µg/mL). Bars represent the standard error of three technical replicates. The median for Col-*0* is from three to four samples per screening assay with three technical replicates each.

3.2.5 Mutant lines 88.06, 89.29, 89.37, 90.32, 90.40, 90.43 and 90.53

3.2.5.1 The insertion site of mutant lines 88.06, 89.29, 89.37, 90.32, 90.40, 90.43 and 90.53 was mapped on chromosome 5

The insertion site of the lines 88.6, 89.29, 89.37, 90.32, 90.40, 90.43 and 90.53 was mapped on chromosome 5, flanking a gene encoding a protein of unknown function. The gene is present in two identical copies. In addition to the gene, the intergenic region is also partly identical. For this reason, the insertion site was allocated to two possible positions. The T-DNA insertion could be either between the two identical genes At5g55880 and At5g55890, or it could flank only gene At5g55880 (Fig. 47).

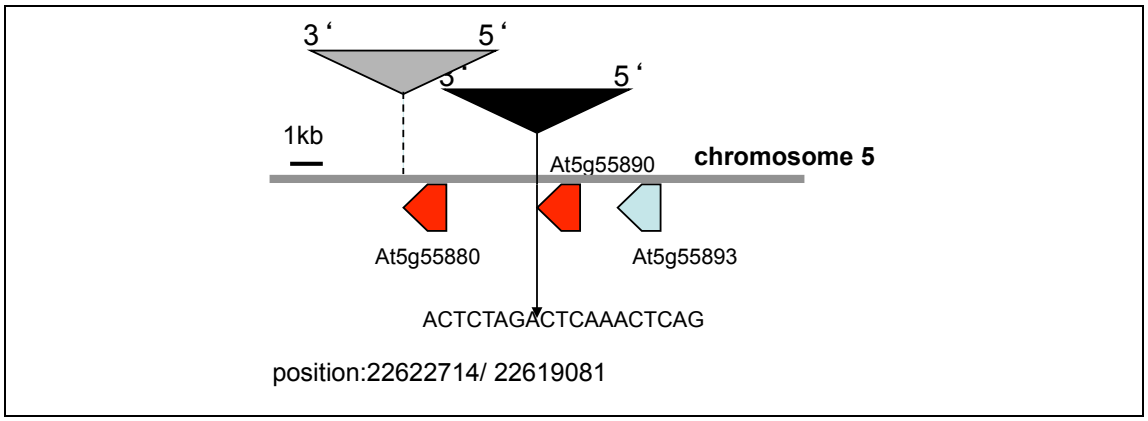


Fig. 47: Insertion site of mutant lines 88.6, 89.29, 89.37, 90.32, 90.40, 90.43 and 90.53.

Expression analyses revealed that the gene(s) was up-regulated up to 100,000 fold. *In silico* analysis using ARAMEMNON uncovered that the genes At5g55880 and At5g55890 are part of a cluster of 27 genes in *A. thaliana* (*159*). None of the members of this gene cluster have been characterized to date. Strikingly, several duplication events have taken place within the cluster. Most of the related genes are located on chromosome 5, ten members can be found on chromosomes 1, 2 and 4. The genes of this cluster have one common domain; the function of this domain is also unknown (Fig. 48). No microarray data is available for the genes in public expression databases.

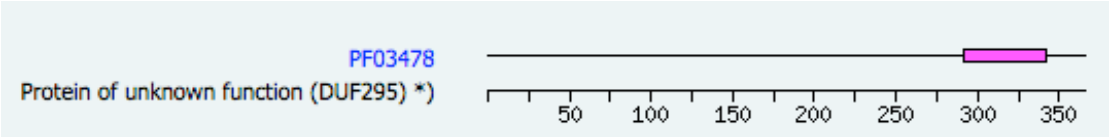


Fig. 48: Protein encoded by At5g55880/At5g55890 with the DUF295 domain (*159*).

3.2.5.2 Subcellular localization via GFP fusion protein

To learn more about the protein encoded by gene At5g55880/At5g55890, the gene was cloned and recombined into the vector pGWB5. Under the control of the CaMV35S-promotor, C-terminal GFP fusion protein was transiently expressed in *A. thaliana* cultured cells and *N. benthamiana* leaves. The GFP fluorescence was observed using a confocal microscope (2.7.2). Both in the cultured cells and in the tobacco leaves, the GFP was visible throughout the cell in small dots (Fig. 50, Fig. 51). *In silico* analysis, using ARAMEMNON, predicts that the protein is localized to the mitochondrion (Fig. 49).

Fig. 49: Predicted subcellular localization of At5g55880 protein using ARAMEMNON (*159*).

The overlay image, which shows the autofluorescence of the chloroplasts in tobacco, reveals that the protein was not localized to the plastids. Together with the *in silico* analysis it can be concluded that the protein is localized to the mitochondrion.

Fig. 50: Confocal microscopy of *A. thaliana* cultured cells transformed with 35S:At5g55880-GFP.

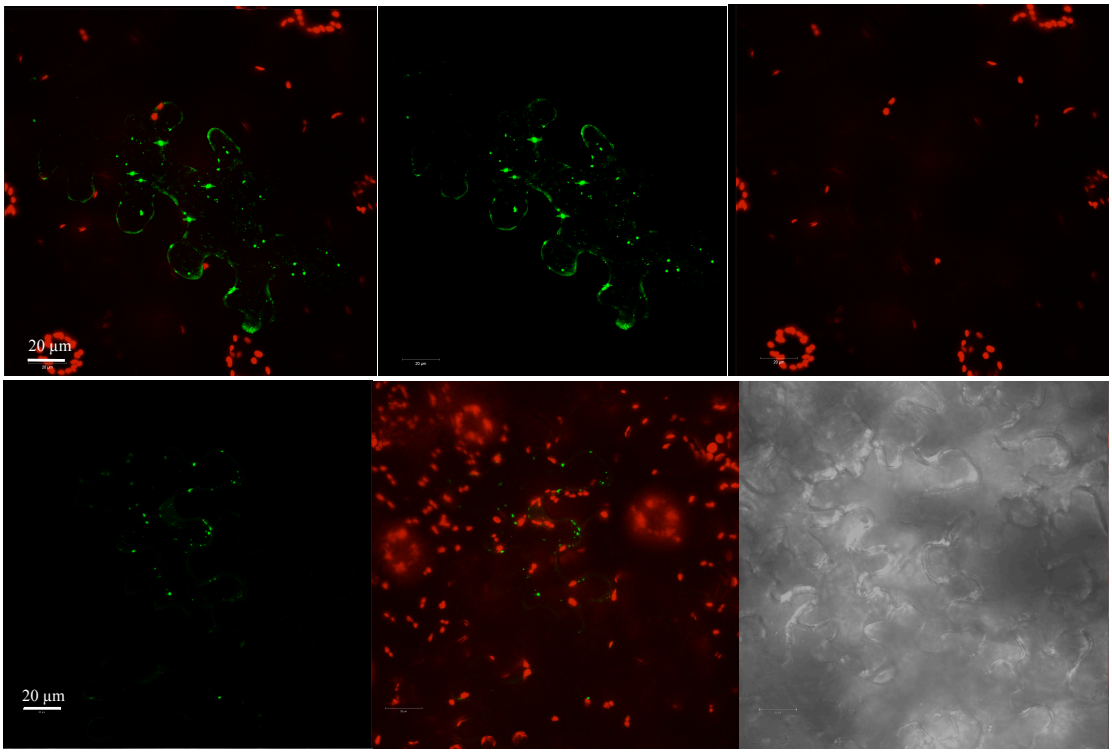


Fig. 51: Confocal microscopy of GFP in *N. benthamiana* leaves transformed with 35S:At5g55880-GFP. The chloroplasts exhibit red autofluorescence. GFP is visible in green.

3.2.5.3 Metabolite profiling of mutant lines 90.32, 90.40 and 90.43

Extracts of the mutant lines 90.32, 90.40 and 90.43 were subjected to metabolite profiling, performed by Dr. Stephan Schmidt (IPB, Halle). The mutants accumulated unknown compounds with long carbon chains with up to 34 carbons (data not shown). While initially it was suggested that these compounds are long-chain fatty acids, subsequent studies did not confirm an accumulation of fatty acids in the mutant lines (personal communication Prof. Dr. Ivo Feußner, University of Göttingen) and future analyses are needed to conclusively find the identity of the accumulated metabolites.

3.2.6 Mutant lines 41.66, 41.16, 41.58, 40.36 and 40.76

3.2.6.1 The insertion site of mutant lines 41.66, 41.16, 41.58, 40.36 and 40.76 was mapped on chromosome 4

During the mapping process, five mutant lines were identified with the same insertion site, namely 41.66, 41.16, 41.58, 40.36 and 40.76. The mutant was named *GH3.5-1D-cp*, since the insertion site was localized on chromosome 4 at position 12649372

between genes At4g27250 and At4g27260. Gene At4g27260 encodes for the IAA amido synthetase AtGH3.5 (Gretchen Hagen) and gene At4g27250 encodes for a putative NAD-dependent epimerase/dehydratase. In the analyzed cDNAs, the GH3.5 gene was up-regulated up to 152 fold, while the expression level of the putative NAD-dependent epimerase/dehydratase was slightly down regulated or remained unchanged. The insertion site of mutant *GH3.5-1D-cp* is shown in Fig. 52.

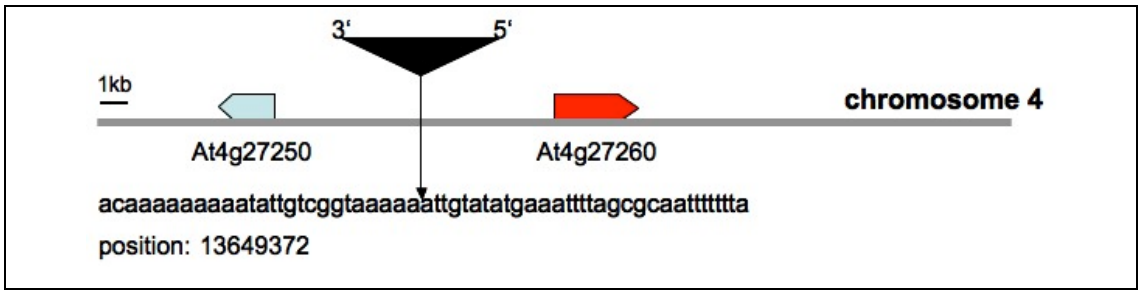


Fig. 52: Insertion site of mutant lines 41.66, 41.16, 41.58, 40.36 and 40.76.

As already mentioned in chapter 3.1.4.2, the mutant *GH3.5-1D-cp* exhibited a strong phenotype. Depending on the expression level of GH3.5, the plants were dwarfish and had curled rosette leaves (Fig. 53). The detected lines had a variable expression level of the GH3.5 gene. The dwarfish phenotype and the curled leaves reflect the typical morphology for auxin imbalance and correspond with the phenotype observed with other activation tagged GH3.5 mutants, namely the *gh3.5-1D* and *wes1-D* mutants (*185,186*). An activation tagged mutant (*dfl1-D*) of the closely related AtGH3.6 has also been detected with a similar phenotype (*187,188*).

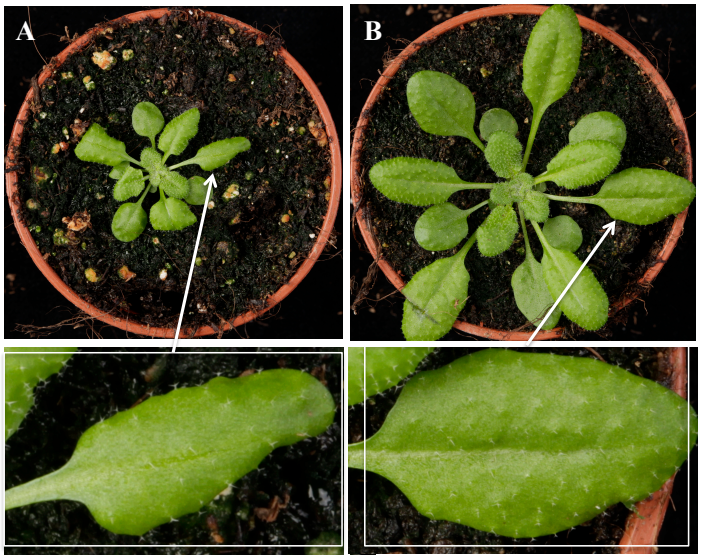


Fig. 53: Phenotype of mutant line (A) 41.66 and (B) Col-*0*. The plants were grown under short day conditions in a growth chamber.

3.2.6.2 Metabolite profiling of *GH3.5-1D-cp*

The mutant *GH3.5-1D-cp* was subjected to UPLC-ESI-QTOF-MS analysis performed by Dr. Stephan Schmidt (Leibniz Institute of Plant Biochemistry). The mutant accumulated elevated amounts of oxindole-3-acetyl-glucose ester/glucoside (OxIAA-Glc) (Fig. 54), the aliphatic glucosinolates glucosibarin (7MSOH) and glucohirsutin (8MSOO), as well as I3M.

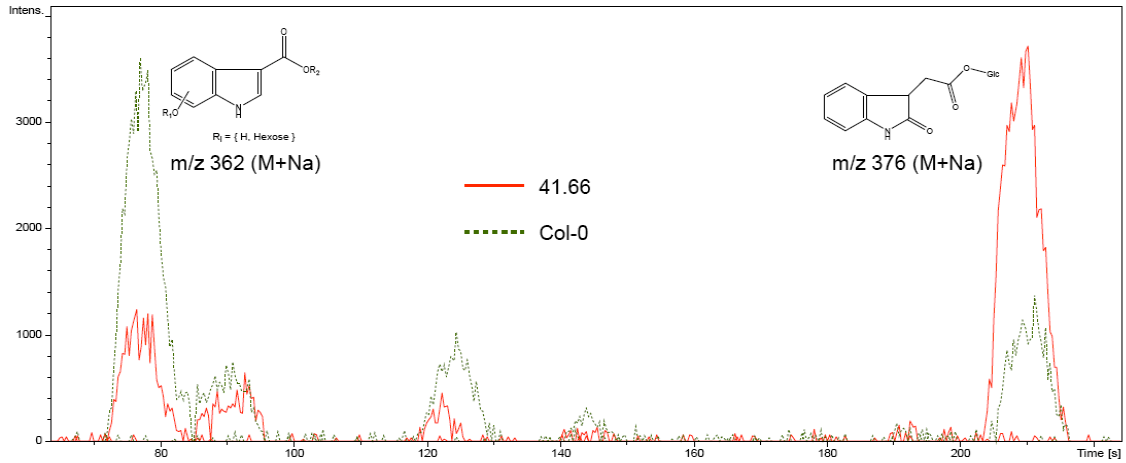


Fig. 54: Excerpt of the chromatogram of mutant *GH3.5-1D-cp*, which accumulated OxIAA-Glc. The chromatogram was provided by Dr. Stephan Schmidt (IPB, Halle).

To confirm the data obtained via UPLC-ESI-QTOF-MS, glucosinolates were extracted from the leaves and analyzed via UPLC as described in chapter 2.7.4. In the

follow-up measurements the glucosinolate content was also altered, the specific pattern observed from UPLC-ESI-QTOF-MS, could not be confirmed though. Other metabolic derangements were suggested such an increase of 8MSOO (experiment 1) and modestly reduced concentrations of I3M (experiments 1 and 2). All in all, the overexpression of GH3.5, resulting in the production of auxin conjugates, leads to an auxin imbalance in the plant. As a side effect, the glucosinolate content seems to be affected. Factors that determined the different glucosinolate content in Halle and Cologne might have been varying expression levels of GH3.5 in the different experiments and the plant cultivation conditions.

3.2.7 Mutant line 68.24

3.2.7.1 The insertion site of mutant 68.24 was mapped on chromosome 1

The T-DNA insertion site of mutant line 68.24 was localized on chromosome 1, at position 26773523 between the genes At1g71000 and At1g71002 (Fig. 55). The gene At1g71000 encodes for an Hsp40/DNAJ-type molecular chaperone. At1g71000 was up-regulated 4.5 fold in the analyzed sample. The gene At1g71002 encodes a microRNA (miR858a). The gene MIR858a was up-regulated up to 1320 fold in the analyzed cDNA samples.

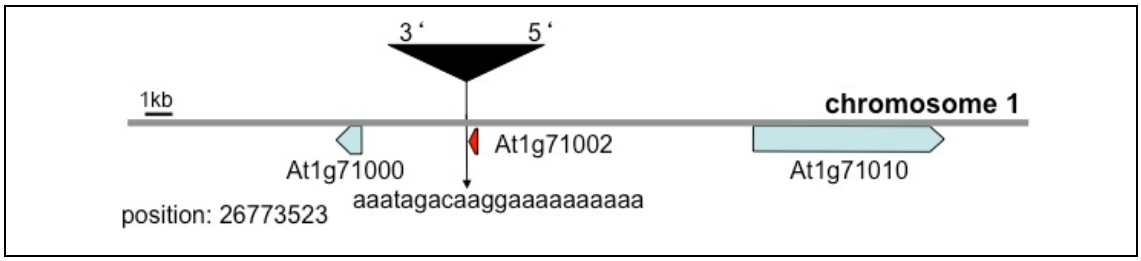


Fig. 55: Insertion site of mutant line 68.24.

3.2.7.2 Putative targets of miR858a and the effect on the phenylpropanoid pathway

On the TAIR database miR858a is annotated to target several MYB transcription factor family members. The mature sequence of miR858a is predicted as follows: UUUCGUUGUCUGUUCGACCUU (*158*). The WMD3 MicroRNA Designer target search tool, predicts 22 putative target MYBs of miR858a: MYB80, MYB12, MYB83, MYB4, MYB19, MYB120, MYB13, MYB71, MYB11, MYB17, MYB20,

MYB42, MYB111, MYB84, MYB37, MYB123, MYB74, MYB87, MYB116, MYB99, MYB6 and MYB85 (*189*). Experimentally, via degradome sequencing, determined targets of miR858a are MYB12, MYB111, MYB13 and MYB20 (*190*). Fahlgren *et al*. (2007) identified MYB12 and MYB83 as putative targets of miR858a via deep sequencing and computational analysis (*191*). The transcription factors MYB11, MYB12 and MYB111 have a regulatory function in the phenylpropanoid pathway. Arabidopsis triple loss of function mutants of MYB11, MYB12 and MYB111 (*myb11/myb12/myb111*) do not produce flavonols in seedlings, while the anthocyanin content is induced (*149*).

Due to the known chemoprotective activity of phenylpropanoid derived compounds, a first experiment was performed testing the expression of MYB11, MYB12 and MYB111 in the available 68.24 cDNA from the screening (experiment performed by L. Marbaise, see (*170*)). The transcription factors MYB51 and MYB28 functioned as controls because they are not predicted targets of miR858a. The high overexpression of MIR858a (1320 fold) resulted in a strong downregulation of MYB11 (0.001 fold), MYB12 (0.019 fold) and MYB111 (0.157 fold), while the expression of MYB28 (0.94 fold) and MYB51 (1.7 fold) remained unaffected. These results suggest that among others MYB11, MYB12 and MYB111 are targets of miR858a. In addition, genes that encode enzymes involved in the phenylpropanoid pathway were tested, namely At5g13930, At3g55120, At3g51240, At5g08640, At5g42800, At2g37040, At3g53260, At5g07990 as well as MYB13 and MYB20, which are two putative targets of miR858a. The results together with the full gene names are listed in Table 13. The chalcone synthase (CHS), chalcone isomerase (CHI), flavanone-3-hydroxylase (F3H), and flavonol synthase (FLS) are transcriptionally controlled by MYB11, MYB12 and MYB111. Except for CHS, which showed a 5.6 fold upregulation, these genes were all downregulated in the tested cDNA. In addition the genes PAL1 (phenylalanine ammonia-lyase) and PAL2, which catalyze the production of cinnamate from phenylalanine were less expressed than in wild-type plants. The effect on the expression of genes F3'H and DFR was only minimal. The putative miR858a target gene MYB13 was downregulated and MYB20 was slightly up-regulated.

Table 13: Relative gene expression of different enzymes involved in the phenylpropanoid pathway.

Gene	Name	Relative gene expression (fold change)
At5g13930	chalcone synthase (CHS)	5.6 x
At3g55120	chalcone isomerase (CHI)	0.139 x
At3g51240	flavanone-3-hydroxylase (F3H)	0.042 x
At5g08640	flavonol synthase (FLS)	0.57 x
At5g42800	dihydroflavonol 4-reductase (DFR)	1.3 x
At2g37040	phenylalanine ammonia-lyase (PAL1)	0.092 x
At3g53260	phenylalanine ammonia-lyase (PAL2)	0.335 x
At5g07990	cytochrome P450 monooxygenase (F3'H)	0.783 x
At1g06180	MYB13	0.256 x
At1g66230	MYB20	1.851 x

Downregulated genes are marked in grey.

To confirm the obtained results, further 68.24 cDNAs were tested. Due to silencing of MIR858a in individual plants of line 68.24, those plants showing an overexpression of MIR858a were selected and the expression of MYB11, MYB12 and MYB111 was determined in each cDNA. An upregulation of the target genes was observed in some of the mutants that were overexpressing MIR858a. It was hypothesized that breakdown products of the target mRNA were amplified. The predicted miRNA target site lies at the 5'-end of the genes and this region is similar in all three genes. In order to distinguish between the genes the applied primers annealed to the 3'-end of the mRNA (cDNA), and did not span the miRNA target site.

To check that the amplified sequences were not cleavage products a semi quantitative PCR was performed using primers that flank the putative cleavage site. This was performed for MYB11 and MYB111, since the expression of MYB12 is mainly in the roots. MYB11 was downregulated in lines 48, 4.3, 42 and 4.1 and could not be detected in lines 43 and 25. MYB111 was expressed in lines 48, 4.3, 42 and 4.1 and was also not detected in lines 43 and 25. It seems that a relatively high expression of the gene MiR858a is necessary in order to reduce the target MYB mRNA.

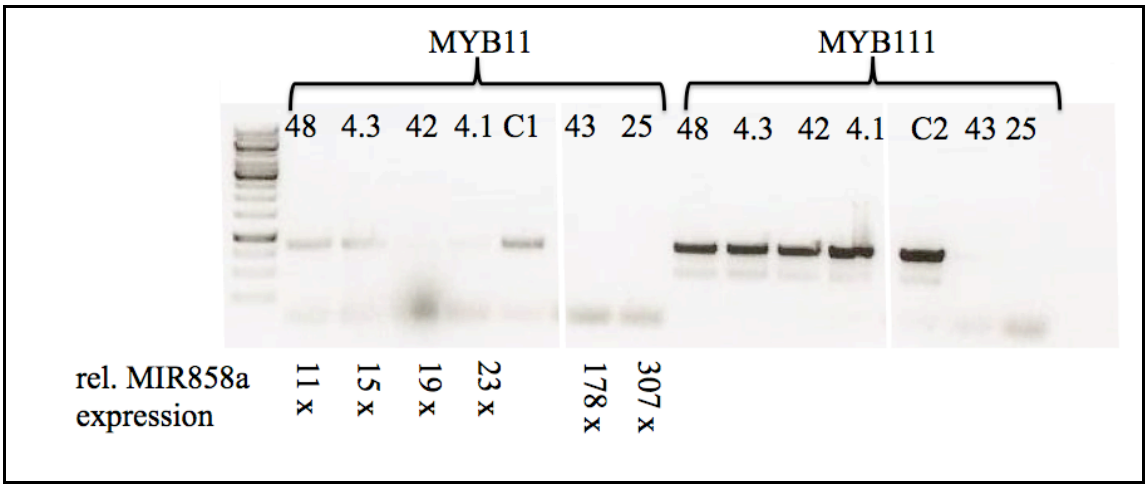


Fig. 56: Agarosegelelectrophoresis of a semiquantitative PCR of genes MYB11 and MYB111 in cDNA of different 68.24 plants. The fragments have sizes of 882 and 912 bp, respectively. The PCR was performed with 39 cycles and 55 °C annealing temperature. C1 and C2 are controls of wild-type plants. The actin ct values were determined via real time PCR: 48 (Ct_{actin}=17.7), 4.3 (Ct_{actin}=19.02), 42 (Ct_{actin}=17.6), 4.1 (Ct_{actin}=17.6), 43 (Ct_{actin}=21.6), 25 (Ct_{actin}= 21.2), C1 (Ct_{actin}= 17.4) and C2 (Ct_{actin}= 17.5).

Taken together, the data suggest that miR858a targets the transcription factors MYB11, MYB12 and MYB111. This influences the expression of genes involved in the phenylpropanoid pathway. In addition further targets of miR858a might exist, such as MYB13. At a relatively high expression of MIR858a (178 fold), MYB11 and MYB111 were downregulated. An overexpression of MIR858a at a lower level (ca. 15 fold for MYB11 and less than 100 fold for MYB111) most likely also led to cleavage of target genes but low levels of target mRNAs were still present.

3.2.7.3 The expression of MiR858a at different day times

The analysis of the MIR858a expression pattern was performed to gain more insight into the role of miR858a in Arabidopsis. One hypothesis is that microRNAs are derived from duplication events of their target gene(s), this can also include the promotor region of the founder gene (*192,193*). In addition, evidence exists that microRNA expression is regulated by the target transcription factors in a negative feedback loop (*194,195*). If the target genes of miR858a are in fact MYB11, MYB12 and MYB111, then one possibility is that these transcription factors also regulate the expression of MiR858a. Therefore, different attempts were made to learn more about the regulation of MIR858a.

From microarray data it is known that the expression of MYB12 changes after light-to-dark transition (personal communication with Dr. Ralf Stracke, University of Bielefeld) therefore the expression pattern of MIR858a during the course of a day was analyzed. In a first experiment the expression levels of MIR858a in leaves of adult Col-*0* plants and plants of line 68.24 were analyzed. Despite a high variance between the expression levels of MIR858a in the individual plants, it became apparent that the expression of MIR858a did not stay constant throughout the day (data not shown). This could also be confirmed for MYB11. In order to reduce the high variation between the samples, Col-*0* seedlings grown on ½ MS agar plates, under short day conditions in the growth cabinet. The 6-day-old seedlings were harvested in pools of ca. 20 individuals. The RNA was extracted using the Z6-extraction protocol (see 2.4.1.3.1.2). The samples were taken at the end of the night (8:45), after the light had been turned on for one hour (11:00), shortly before the light would turn off (17:30) and 1½ h after start of the night (19:30). Expression analyses revealed that the expression of MIR858a was the highest at the end of the day (17:30), while it stayed relatively constant at the end of the night and start of the day (8:45, 11:00). After the night had started, the expression of MIR858a was down regulated again (Fig. 57).

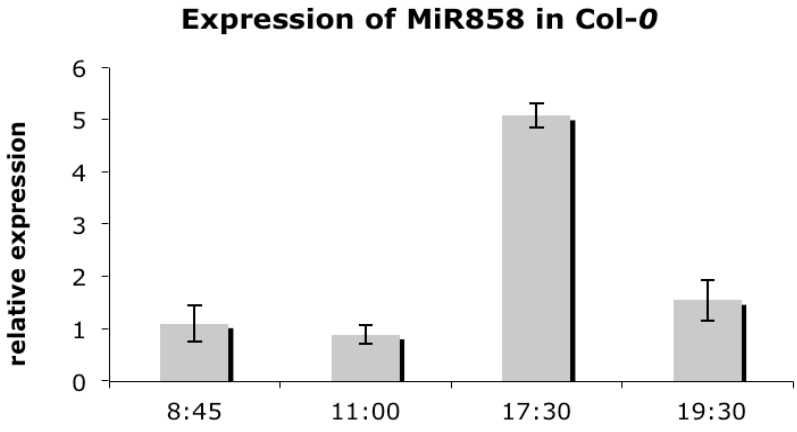


Fig. 57: Relative expression of MIR858a in pooled Col-*0* seedlings depending on daytime. The relative expression was calculated in relation to the average expression at 8:45. RNA samples were taken from ca. 20-pooled seedlings 6 days of age. The plants were grown on ½ MS agarose plates under short day conditions. The relative expression values are an average of three pools of one biological replicate. The bars represent the standard error. The repetition of the experiment showed similar results.

In a further experiment, the expression of MIR858a in the triple loss of function mutants *myb11/myb12/myb111* was tested throughout the day (the mutants were kindly provided by Dr. Ralf Stracke, University of Bielefeld). The expression of MIR858a in these plants varied at the different day times and did not follow the pattern observed in Col-*0* wild-type plants. The results suggest that the expression of MIR858a is influenced by the putative target transcription.

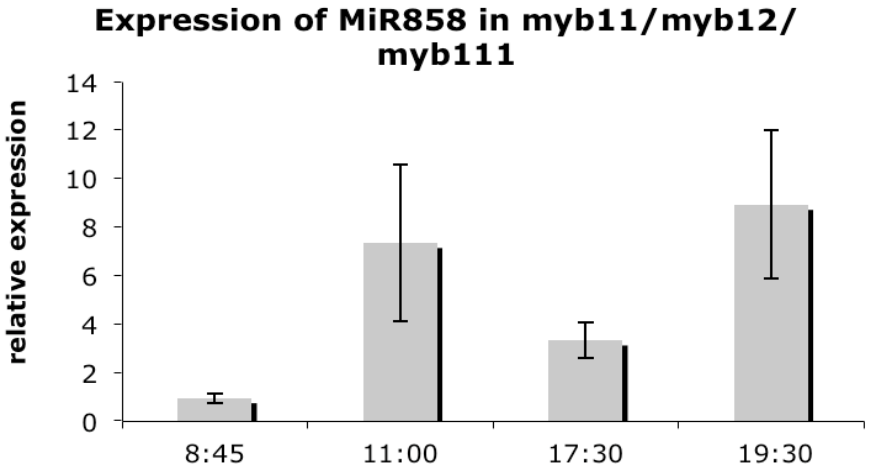


Fig. 58: Relative expression of MIR858a in pooled *myb11/myb12/myb111* triple knockout mutant seedlings depending on daytime. The relative expression was calculated in relation to the average expression at 8:45 in Col-*0* plants. RNA samples were taken from ca. 20-pooled seedlings 6 days of age. The plants were grown on ½ MS agarose plates under short day conditions. The relative expression values are an average of three pools. The bars represent the standard error.

3.2.7.4 Analysis of MIR858a expression in different tissues

To learn more about the expression pattern of MIR858a within the plant, the GUS gene was placed under the control of the putative promotor region of MIR858a, using the vector pGWB3i. In general the promotor region of MIR genes is predicted within the upstream region of 800 nt (*194*). A region of 2082 nt, including the MIR858a gene sequence, was cloned. Different organs of the T_1-generation were tested for GUS activity. The promotor of MIR858a was active in the inflorescences, the veins of rosette leaves and in the stems of the siliques, but not in the siliques and seeds (Fig. 59).

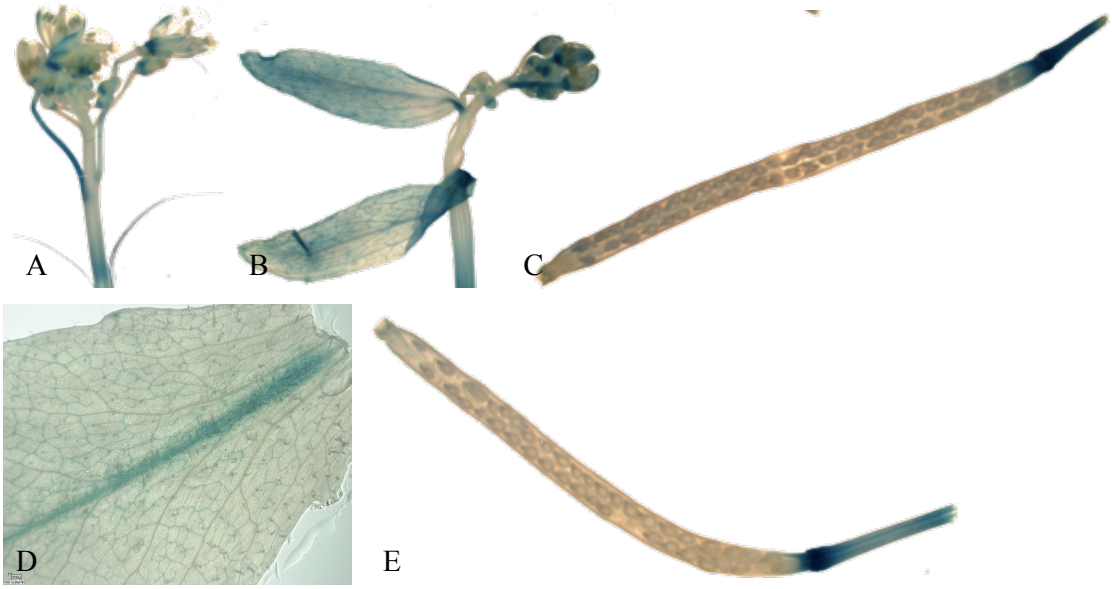


Fig. 59: GUS activity in inflorescences (A,B), siliques (C,E) and leaves (D) of PromMIR858a:GUS lines. The GUS activity was observed in different lines of the T_1-generation of transformants. Representative pictures from different lines are shown here.

The promotor activity was also analyzed in developing seedlings from two-day-old to eight-day-old plants of different lines. The samples were always taken at the same daytime on each day. From the staining it becomes apparent that MiR858a is expressed during seedling development. After germination, the promotor was active at the tip of the radicle and in the cotyledons, which were still inside the seed coat. On day three, the seedling had shed the seed coat and GUS staining was visible in the vascular of the cotyledons and in the hypocotyl and in the tip of the root (meristem). On days four and five the GUS distribution was similar as on day three. The eight-day-old seedlings exhibited GUS activity in the vascular of the leaves and in the hypocotyl. There was also intensive GUS staining visible at the hypocotyl-root junction (Fig. 60).

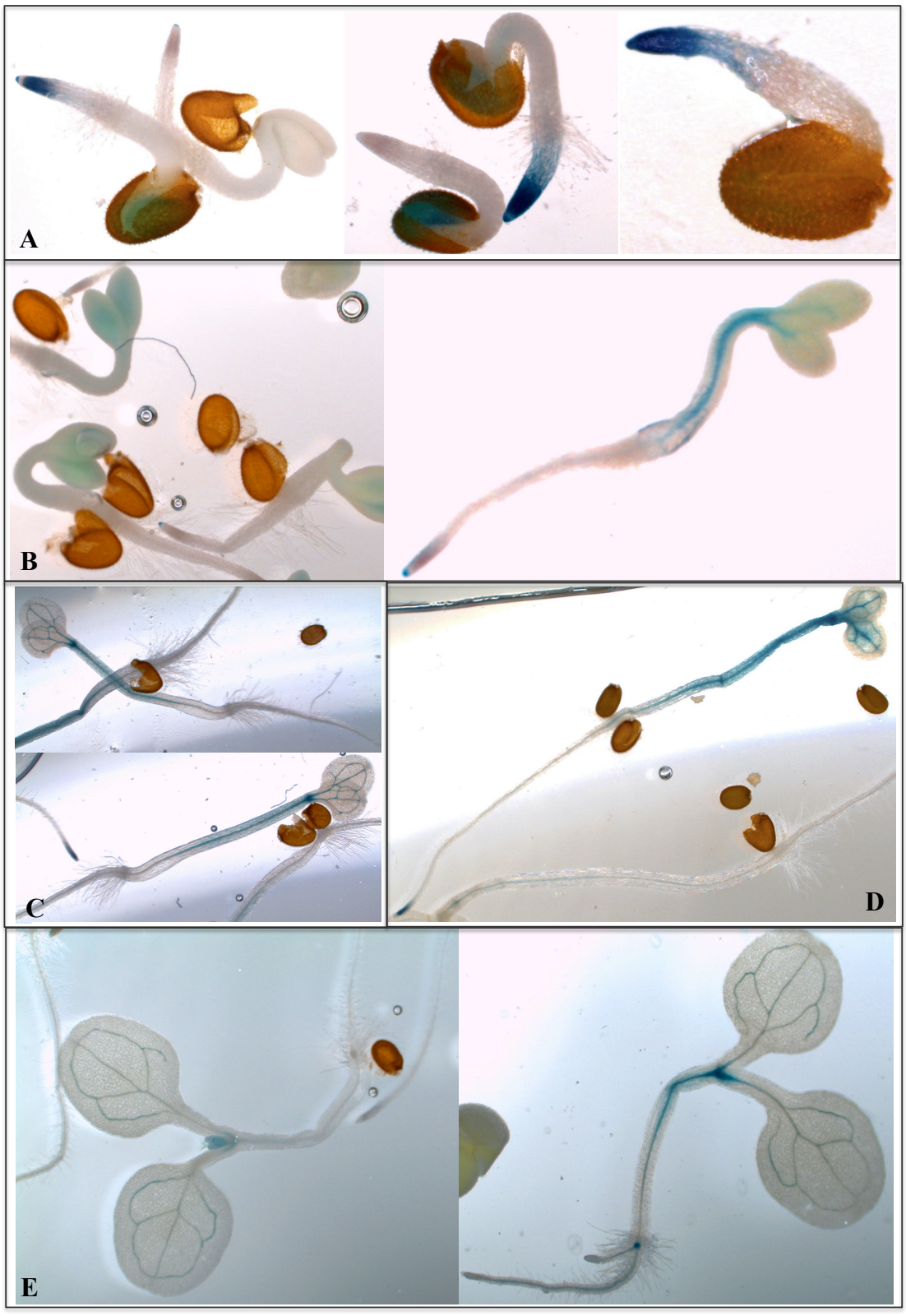


Fig. 60: MiR858a promotor GUS analyses in two to eight-day-old seedlings. (A) two-day-old seedlings (lines V80.3, 16a, 6a) (B) three-day-old seedlings (lines V80.3, V80.6) (C) four-day-old seedlings (lines V80.3, V80.6a) (D) five-day-old seedling (lineV80.6a) (E) eight-day-old seedlings (lines V80.1a, V8016c). The seedlings were grown on ½ MS-medium under short day conditions. T_2 generation plants were analyzed.

It could be shown that MYB12 and MYB111 are regulated by the bZIP transcription factor HY5 in response to light and UV-B radiation (*196*). Light regulatory elements were also detected in the promotor region of MIR858a (*197*). Here the effect of different light intensities was tested using the promotor-GUS lines. The seedlings were transferred from short day conditions to long day and high light conditions. After one week, the GUS activity was determined. The GUS assay was performed with 13-day-old seedlings of two independent lines. Under short day conditions the staining was limited to the shoot apex region and the base of the leaf stems. A very mild staining was visible in the vascular tissue of the leaves. Seedlings that were transferred to the long day exhibited stronger staining. The staining was visible in the shoot apical meristem (SAM), the leaf stems and very mildly in the roots. Under high light conditions the GUS activity was the highest in both lines. It was clearly visible in the vascular tissue of the leaves, the SAM and the roots. In summary the data suggest that the promotor of MIR858a is induced upon increased intensity of light (long day and high light).

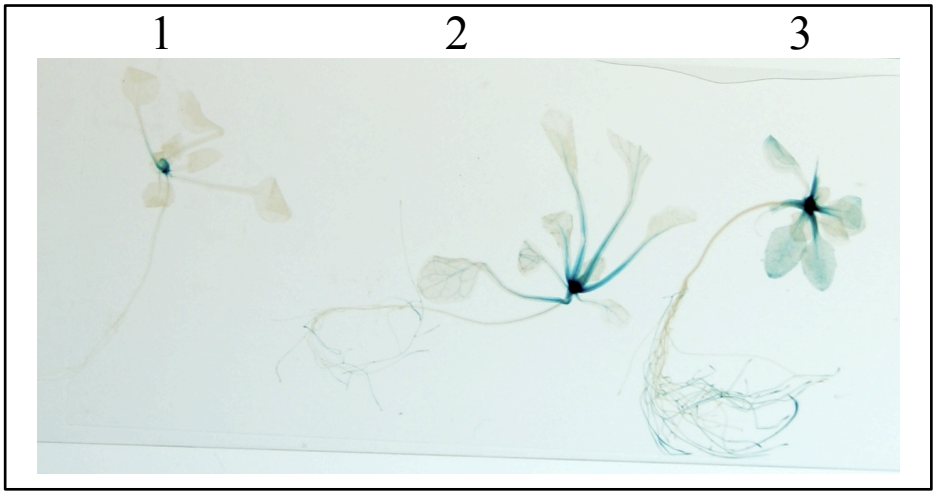


Fig. 61: Treatment of promMIR858a:GUS seedlings with different light conditions. After the seedlings were grown on ½ MS-medium for 6 days under short day conditions, they were transferred to (2) long day, (3) high light or (1) remained at short day. After 7 days the GUS activity was determined.

3.2.7.5 MIR858a mutant analysis

To confirm the gene-to-trait relationship the MIR858a gene was cloned into the expression vector pGWB2 and Col-*0* plants were transformed with the construct. A number of resistant T_1-lines were isolated but subsequent expression analyses via real time PCR revealed that none of the lines overexpress MIR858a more than ca. 10-fold

of wild-type levels. Nevertheless for an initial experiment with Hepa1c1c7-Lux cells, extracts of line 35S:MIR858a9 were tested. In addition plants overexpressing a miR858 target mimic (MIM858) were cultivated (MIM858 lines were kindly provided by I. Rubio-Somoza, MPI for Developmental Biology, Tübingen). Since classical loss-of-function mutants via T-DNA insertion are not available, these plants were constructed to inhibit the function of a specific miRNA (*150*). In one experiment, extracts of MIM858, MIR858a and 68.24 induced the reporter cells. This was not the case in a second screen. Since also extracts of 68.24 failed to induce luminescence in this screen, the lack of induction was probably due to a problem with the reporter cells.

4 Discussion

4.1 Screening of TAMARA mutants with the Hepa1c1c7-Lux reporter cells

The anticancerogenic effect of many phytochemicals can be ascribed to their induction of phase II and detoxification enzymes. Some phytochemicals can additionally inactivate carcinogens by directly functioning as anti-oxidants (*110*). To test the chemoprotective activity of a compound, the expression of phase II and detoxification enzymes and the activation of Nrf2 in response to treatment with the respective compounds or extracts were analyzed (*5,198-200*). To simplify these tests, reporter systems have been developed in which a reporter gene is placed under the control of EpRE. The HepG2-GFP and Hepa1c1c7-Lux reporter cells developed by Zhu *et al.* (2000) and Boerboom *et al.* (2005) are such systems (*112,113*). EpRE based reporter systems have been applied to test single compounds and extracts from different plants (*201-209*). In this way the molecular characteristics of a chemoprotective compound could be eluciated such as the role of planarity of flavonoid molecules in the induction of phase II enzymes (*113*).

Here, the EpRE based reporter systems were applied for testing the chemoprotective activity of compounds found in *A. thaliana* Col-*0* plant extracts. In contrast to using isolated substances, plant extracts contain a variety of different compounds. These compounds may have inducing, inhibiting or even toxic effects on the cells. When administered in a high concentration the compounds may be toxic (*210*). This was observed with extracts from grape and *A. thaliana* (Col-*0*) (*114,208*). Therefore, the assessment of an appropriate amount of Col-*0* extract was a first necessary step in this work. The fact that the secondary metabolism of a plant is easily affected when the environmental conditions change is a further challenge (*211*). It was therefore important to cultivate the plants under controlled conditions. A treatment of the Hepa1c1c7-Lux reporter cells with Col-*0* extracts led to a basal induction of luciferase due to the high amount of phase II enzyme inducers found in this plant. 80% of the induction activity of the extracts can be ascribed to methylsulfinylalkyl glucosinolates, mainly glucoraphanin (*114*).

The EpRE based cell culture reporter system is a valuable tool for the discovery and analysis of chemoprotective compounds. To test the Hepa1c1c7-Lux cells, extract of the *HIG1-1D* mutant was used as an inducer. The overexpression of MYB51 in this mutant leads to the accumulation of I3M (*151*). The breakdown product I3C is a known chemoprotective compound (*61,166*). The mutant extracts led to an increased luciferase signal. Hepa1c1c7-Lux cells thus enable the screening of *A. thaliana* plant extracts.

As with most model systems, some limitations though exist with using cell models. One aspect is the fact that the cells with which we are working are cancer cells. Many aberrations have taken place in these cells. Some phytochemicals even lead to apoptosis of cancer cells (*212*). The cancer cells have lost many characteristics such as contact inhibition. In addition the cells are not in their *in vivo* environment and in contact with other cell types from different tissues. Many of the compounds consumed by mammals will not reach the organisms' systemic circulation in their original state after they have passed the digestive tract. Therefore, the reporter cells treated with plant extracts are exposed to compounds and compound levels, which they would not encounter *in vivo* (*213*). Recently, it was demonstrated that increased constitutive Nrf2 transcription may even promote tumorigenesis in the analyzed cells (*214*). This demonstrates that the use of EpRE based reporter systems for the search of new phytochemicals is only one aspect of chemoprevention and that many other factors play a role in cancer development and intervention. Nevertheless, the cell culture reporter systems offer the opportunity to discover new phytochemicals in a high-throughput and ethical manner. The discovered compounds can later be tested in animal models and clinical studies. In addition, the screening allows the further elucidation of secondary metabolism in plants.

4.2 Analysis of the discovered genes

Approximately 5,000 TAMARA mutants were screened using the HepG2-GFP cells with the goal of finding new chemoprotective compounds and genes involved in the metabolite production. After two rounds of screening, ca. 470 candidates were selected for further analyses using Hepa1c1c7-Lux cells. 149 mutants were mapped, which led to the identification of 87 different mutants. The expression of the genes ultimately up and downstream of the insertion site was determined. These data give an

indication of which genes may be responsible for the EpRE inducing effect of the mutant extract. In order to get a general picture, 158 discovered genes were categorized into different groups. In a first attempt, the groups were formed on the basis of the putative or known gene functions listed in TAIR and ARAMEMNON. 32% of the genes encode for enzymes, 13% for transcription factors, 11% are involved in transport, 26% are of unknown function and 18% could not be categorized into a general group.

The drawback of the clustering or grouping of those genes showing an altered expression level is the fact that in many mutants more than one gene had an altered expression. The chemoprotective effect may be allocated to only one of the genes. Since at this point it can only be speculated which of the genes is responsible for the chemoprotective activity of a mutant, the analysis of all altered genes can be misleading. On the other hand the chemoprotective effect may also be due to the altered expression of a combination of genes. Neighboring genes can for examples have similar functions due to duplication events.

With the help of the bioinformatic tool DAVID the genes detected during the screening can be sorted on the basis of functional annotations. This form of clustering gives more detailed insight to the functions of the genes such as the pathway in which it may be involved. The analysis resulted in 25 clusters. Some of their members overlap with members of other clusters.

The discovered groups include enzymes involved in macromolecule catabolic processes, mainly glycosyl hydrolases of the family 1, 9, 19 and 28. The majority of these glycosyl hydrolases are predicted to function as β-glucanases, chitinases and polygalacturonases that are associated with cell wall organization, which also reflects the situation in the plant, where most glycosyl hydrolases are involved in cell wall polysaccharide metabolism (*215*). Some of the detected genes in this group respond to abiotic or biotic factors, such as water, light and fungal pathogens. Members of the class of glycosyl hydrolases have been reported to play a role in the metabolism of cell wall polysaccharides, biosynthesis and remodulation of glycans, mobilisation of storage reserves, plant defense, symbiosis, secondary metabolism, glycolipid metabolism and signaling (*215*). Glycosyl hydrolases can be directly involved in the production of chemoprotective secondary compounds via the hydrolysis of secondary metabolites, which influences their biological activity, chemical stability and water

solubility. The activation and inactivation of phytohormones, such as auxins, cytokinins, gibberellins, abscisic acid and SA can also be caused via hydrolysis by glycosyl hydrolases. In this way glycosyl hydrolases can affect signaling in the plant which can result in changes of secondary metabolism (*215*).

The second category of genes is involved in transport. Members of this group have transmembrane domains. Representatives of both passive and active transport were found. The detected transporters include among others several putative aquaporins and two ABC transporters. The altered expression of plasma membrane transporters can have an effect on nutrient availability in the plant and therefore can have an impact on plant growth and metabolism. An example is the overexpression of the aquaporin PIP1b, which leads to increased plant growth rate, transpiration rate, stomatal density, and photosynthetic efficiency under favorable growth conditions (*216*). In addition, some transporters have been reported to transport phytohormones such as amino acid transporters of the amino acid/ auxin permease family for auxin transport or purine permeases for cytokinins (*217*). The altered transport of phytohormones effects plant signaling and can impact secondary metabolism. Transporters of secondary metabolites have also been identified, which offer a target for increasing the production of secondary metabolites in addition to altering the expression of biosynthetic enzymes (*218,219*).

The third group consists of genes involved in reproductive organ development such as the floral organs and the seeds and post-embryonic organ development. This group of ten members contains several transcription factors, namely WRKY35, MYB26, WOX2, Brevis Radix- type transcription factor, sterile apelata protein transcription factor, and the leafy cotyledon 2 family B3 type transcription factor. Transcription factors have key regulatory roles in plant metabolic pathways and can control multiple steps. Therefore, they are considered as promising factors for plant metabolic engineering. Many transcription factors that regulate plant secondary metabolism are still unknown (*220*). The Brevis Radix-2 is known to be involved in cell proliferation and elongation of the root tip (*221*), mediates the interaction between the brassinosteroid pathway and auxin signaling in roots (*222*) and is involved in cytokinin-mediated inhibition of lateral rot initiation (*223*).

The screening detected several genes involved in secondary metabolic processes. The six members include proteins involved in glutathione metabolism, namely a gamma-

glutamyl transpeptidase, which degrades the anti-oxidant glutathione, and two glutathione-S-transferases. Three proteins were predicted to be involved in the phenylpropanoid pathway. These are two acyl-CoA synthetases and a laccase that is involved in lignin catabolic processes. The acyl-CoA synthetases, At1g20500 and At1g20510 are closely related to 4-coumarate:CoA ligases, which catalyze the formation of CoA-esters from 4-coumarate and cinnamic acid derivatives, building blocks of the phenylpropanoid metabolism. The acyl CoA synthetases found here have been shown to be involved in jasmonic acid biosynthesis. At1g20500 catalyzes the conversion of OPC-6 to OPC-6-CoA ester and can also activate the cinnamic acid derivative sinapic acid. At1g20510 catalyzes the conversion of 12-oxo-phytodienoic acid, (3-oxo-2(2'-[*Z*]-pentenyl)cyclo-pentane-1-octanoic acid) OPC-8, OPC-6 and OPC-4 to the respective esters. These steps are necessary for the β-oxidative shortening of the side chain, which leads to jasmonoyl-CoA formation, a precursor of jasmonic acid (*224*). As mentioned in chapter 1.1.7, chemoprotective activity of jasmonic acid has been demonstrated. Glutathione-*S*-transferases are involved in the detoxification of toxins via conjugation with glutathione. Some form a complex with flavonoids to facilitate the transport into the vacuole, they play a role in stress metabolism and function in stress signaling by inducing chalcone synthase upon exposure to UV-light (*225*). Via the action of glutathione-*S*-transferases putative chemoprotective compounds could be produced.

The fifth group consists of genes that respond to abiotic stresses. As stress response, plants produce a variety of secondary metabolites (*211*). Therefore, genes of this group represent putative candidates for the production of chemoprotective compounds. The eight members forming this group are very diverse. Some of them have already been discussed, such as a glutathione-*S*-transferase (At1g02930) or the plasma membrane intrinsic protein (At1g01620). Also the β-glucosidase PYK10 will be discussed in the following chapter. One group of genes responds to water deprivation, namely the class-III peroxidase (At1g05260), the plasma membrane intrinsic protein and the glutathione-*S*-transferase. The peroxidase, a protein kinase, the glutathione-*S*-transferase and PYK10 respond to salt and osmotic stress and the membrane-extrinsic subunit of photosystem II complex (At5g66570), a fatty acyl-CoA reductase (At5g22500) and a calcium cation permeable channel (At3g04110) respond to light. Members of this group may have signaling functions in stress

response. For example, the protein kinase SnRK2.7 of the sucrose nonfermenting 1-related protein kinase 2 family is activated both by hyperosmotic stress and abscisic acid (ABA). This protein kinase family is involved in osmotic signaling (*226*). ABA and osmotic stress leads to the production of various secondary compounds in plant cells, such as indole alkaloids in cell culture of *C.roseus*, taxol in *Taxus spp.* cell culture, hydroxycinnamic acid amides and tryptophan in barley. ABA inhibits flavonoid accumulation in soybean (*227*).

Another group of genes respond to hormone signaling. The fact that more than half of the ten members of this group encode for transcription factors shows that genes were detected with key regulatory roles in plant signaling. Three members of this group respond to ethylene, namely the ethylene-responsive binding factor (AtERF14), and two putative ERF-type transcription factors (At1g77200, At5g13330). Ethylene is an important factor in growth, development and plant defense and can be induced by wounding, ozone, pathogens and insects. The accumulation of ethylene can have an effect on secondary metabolism. For example in grape cell culture, ethylene treatment increased flavonoid, stilbenoid and anthocyanin production (*227,228*). AtERF14 plays an important role in regulating plant defense responses (*229*).

The group of proteins that bind to metal ions consists of 25 members and is the largest group of genes detected here. In general, metal ions can function as protein cofactors. Metal ion binding to proteins can for example influence the interaction of the protein to DNA (*230*). The diverse group of metal ion binding proteins detected here consists mainly of enzymes. The important role of enzymes in the production of secondary metabolites was discussed already.

Finally, a small group with four members was detected that have F-box domains. F-box proteins function in protein ubiquitination and degradation and nearly 700 F-box proteins are predicted to exist in *A. thaliana* (*231*). The function of the detected genes is still unknown. The genes At1g47810 and At1g70970 are part of a common cluster of unknown proteins (*159*). The gene At3g59590 is related to a myrosinase binding protein. The homology may point to a putative role of At3g59590 protein in glucosinolate hydrolysis via myrosinases. In general F-box proteins have a number of different functions in plants, they function in hormone signaling, such as the receptor TIR1 that binds to auxin, but also play a role in jasmonate, gibberellin and ethylene pathways. F-box proteins are involved in light signaling and the circadian clock,

pollen recognition and rejection, different developmental processes such as lateral root formation and shoot branching, have functions in cell cycle, and defense response (*231-233*).

Examples for each category could be found that are known or predicted to play a role in secondary metabolism. This demonstrates the potential of the detected mutants in the analysis of phytochemical production in plants.

4.3 Analysis of selected mutants

4.3.1 Mutant line *PYK10-1D-cp*

PYK10 is a root and hypocotyl specific enzyme that has been associated with glucosinolate breakdown in roots (*22*). In *PYK10-1D-cp* the transcript level of PYK10 was strongly enhanced in the leaves. Interestingly, PYK10 expression has so far been believed to be limited to the root and hypocotyl of seedlings and was not thought to take place in rosette leaves (*172*). The fact that PYK10 transcript was highly abundant in rosette leaves of *PYK10-1D-cp* demonstrates that PYK10 expression must take place at least in low basal levels in leaves of wild-type plants. Since the breakdown products of glucosinolates are known inducers of phase II and detoxification enzymes, the detection of the *PYK10-1D-cp* line offers new opportunities to study the chemoprotective effects of compounds derived from glucosinolate breakdown.

Until a few years ago, the hydrolysis of glucosinolates was believed to take place only upon tissue disruption. Bednarek *et al.* (2009) discovered that the atypical myrosinase PEN2 degrades the indole glucosinolates I3M and 4MO-I3M to raphanusamic acid and indol-3-ylmethylamin or 4-methoxyindol-3-ylmethylamin in intact cells in response to fungal infection (*36*). Parallels between the PEN2 and PYK10 pathways can help to further elucidate the role of PYK10 in roots. The homologous enzymes are part of a subclade of β-glucosidases and are considered to be atypical myrosinases. The PYK10 coexpression network reveals that the monooxygenase CYP81F4, necessary for the synthesis of 1MO-I3M, is coexpressed with PYK10, so that 1MO-I3M might serve as a substrate of the enzyme (*177*). Moreover, PYK10 is coexpressed with the transcription factor MYB34, which has recently been demonstrated to regulate the production of 1MO-I3M in roots (*178*).

PYK10 is the main component of constitutive ER bodies (*179,234-236*). The transient expression of PYK10-GFP protein in *A. thaliana* cultured cells confirmed the functionality of the clone and the localization of PYK10 protein to the ER bodies (see chapter 3.2.2.2). ER bodies are grouped into the constitutive and inducible. The constitutive ER bodies are present in the roots, hypocotyl and cotyledons of Arabidopsis seedlings (*234,237,238*). ER bodies have been suggested to play a role in plant defense reactions, because they are induced upon wounding and jasmonate treatment (*235,238*).

The mutant lines *PYK10-1D-cp*, *pyk10-1*, and *35S:PYK10* were analyzed for 1MO-I3M, 1MO-ascorbigen, 1MO-IAN, 4MO-I3M, 4MO-ascorbigen, 4MO-IAN and scopolin content in comparison to Col-*0*. The metabolite profiling revealed that the nitriles 1MO-IAN and 4MO-IAN were strongly reduced in the PYK10 loss-of-function mutant. The content of 1MO-I3M was reduced in PYK10 overexpressing lines. Ascorbigen substitutions, namely 1MO-ascorbigen and 4MO-ascorbigen were reduced in *PYK10-1D-cp* and *pyk10-1 lines* but not in the two *35S:PYK10* lines, therefore not giving a clear picture for interpretation. Since PYK10 was found to hydrolyze the compounds scopolin and esculin *in vitro*, the scopolin content was also analyzed in the mutants (*173,174*). No clear results could be obtained from the present data, since the scopolin concentration was almost the same in all mutants but reduced in comparison to Col-*0*. The obtained data on nitrile accumulation suggest that PYK10 is involved in the formation of nitriles in roots. In addition, the reduced amount of 1MO-I3M in the overexpressing lines indicates that 1MO-I3M is a substrate of PYK10. Since 4MO-IAN was strongly reduced in *pyk10-1* it can be concluded that PYK10 is also involved in the formation of 4MO-I3M derived nitriles. The 4MO-I3M content was not affected by PYK10 overexpression, suggesting that this glucosinolate is not a primary substrate of PYK10. PYK10 interacts with a number of other proteins, such as PYK10-binding protein 1 (PBP1) (*239*). In intact cells PYK10 and PBP1 are stored in separate compartments (*239*). For its activation, the enzyme needs to polymerize, making the protein non-soluble. This activation is induced by the PBP1 (*239*). In addition, PYK10 is coexpressed with a number of JALs (jacalin-related lectin) and GLLs (GDSL lipase-like protein). The expression of PYK10, JAL22, JAL23, JAL31, JAL33, PBP1, JAL30, GLL23 and GLL25 is regulated by the basic helix-loop-helix type transcription factor NAI1 (*173,180,239*).

PYK10 forms large complexes with JALs and GLLs (*180*). It is assumed that JALs and GLLs regulate the size of the PYK10 complex and its substrate specificity (*180,237*). The presence of nitrile specifier proteins during hydrolysis of glucosinolates leads to the production of simple nitriles (*27*). JALs or other specifier proteins could regulate the outcome of glucosinolate hydrolysis via PYK10 in a similar fashion. The nitrile content was not higher in overexpressing lines suggesting that this reaction may be dependent on other proteins not up-regulated in the mutants. It could also be the case that nitriles are not the main hydrolysis products of indole glucosinolates catalyzed by PYK10. Nevertheless, presently unidentified compounds may be accumulating in the PYK10 overexpressing lines causing the induction of the reporter systems. From a chemoprotective point of view, it is still unclear, if the formation of nitriles at the cost of other breakdown products is an unwanted effect (*26*). It could be shown though that carcinogens are formed when IAN reacts with nitrite (*60*). In the *pyk10-1* lines, less nitriles were formed and since 1MO-I3M and 4MO-I3M concentrations did not increase, other chemoprotective breakdown products might be produced. Further untargeted metabolic profiling attempts will be necessary to search for accumulating compounds in the overexpressing PYK10 lines. Here it will be necessary to analyze leaf material, since this was the source for the plant extracts during the screening. To find out if these compounds have a function in plant defense it would be interesting to test the resistance of *PYK10-1D-cp* and *35S:PYK10* against different pathogens. Another question to be answered in the future is whether PYK10 protein has a function in rosette leaves. In addition, PYK10 protein will need to be isolated and the substrate specificity for indole glucosinolates further assessed.

4.3.2 Mutant line *AAT-1D-cp*

In the mutant line 52.18 (*AAT-1D-cp*) the transcript level of the putative amino acid transporter At3g54830 was 80-fold higher in comparison to Col-*0*. The gene At3g54830 encodes for a putative amino acid transporter, which belongs to the amino acid/auxin permease family (transporter classification 2.A.18). Due to the identification of elevated amounts of several aliphatic ITCs via metabolite profiling, the mutant was chosen for further studies. The discovered ITCs included 4-MSOB-

ITC, 5-MSOP-ITC, 7-MSOH-ITC and 8-MSOO-ITC. In addition, the indole glucosinolate I3M and its breakdown product ascorbigen also accumulated in higher concentrations. Glucobrassicin derivatives are known inducers of phase II enzymes (*166*). 7-MSOH-ITC and 8-MSOO-ITC are inducers of the phase II enzyme quinone reductase in Hepa1c1c7 cells (*240*). Sulforaphane (4-MSOB-ITC) is derived from the glucosinolate glucoraphanin and is known as a strong inducer of phase II enzymes (*49,52,114,241*). The flavonol kaempferol, which also accumulated in the mutant, is also a chemoprotective compound (*88,89,242*).

Analyses of At3g54830 transcripts revealed that the gene exists in at least three different splicing variants. N-terminal GFP fusion constructs with all three splicing variants were made and a GFP signal for two of the three constructs (clone 5 and clone 2.9) could be detected. The protein was localized to the endomembrane system outside of the vacuole, similar results were found for the related amino acid transporters At5g02180 and At5g65990 of the same family (*181*).

The promotor GUS analysis revealed promotor activity in the vascular tissue of rosette leaves, and in the cotyledons, young leaves and root tips of seedlings. The amino acid transporter At2g39130 is with 66% protein sequence identity the closest relative of At3g54830. The promotor of this gene was active in cotyledons and young leaves, as well as in the roots of developing seedlings and the lateral roots of adult plants. Therefore it was hypothesized that At2g39130 is involved in the supply of nitrogen to sink organs in seedlings and uptake of amino acids from soil. At2g39130 was also expressed in the pollen but this was not analyzed here. Since At3g54830 shows a similar expression pattern as At2g39130, the two transporters may have similar roles (*181*).

In a microarray experiment performed by Schmidt (2007), the amino acid transporter was down regulated upon treatment with phytoprostanes (*243*). These compounds are formed non-enzymatically from α-linolenic acid and are produced during oxidative stress in higher plants (*244,245*). In a preliminary wounding experiment of Col-*0* adult plants and treatment of Col-*0* seedlings with jasmonate and SA, transcript of At3g54830 was also reduced (data not shown). Taken together these data implicate an indirect involvement of At3g54830 in stress response. Upon pathogen infection plants induce defense responses, which can reduce the metabolism of the plant (*246*). Further data will need to be gathered here.

In summary, a direct link between the overexpression of At3g54830 and the accumulation of glucobrassicin and aliphatic ITCs cannot be made at this point. Since especially ITCs are strong inducers of phase II enzymes, they are probably the causal agent of the EpRe induction in HepG2-GFP cells. The induction of the Hepa1c1c7-Lux cells was less convincing as during previous screens with HepG2-GFP cells. A preliminary experiment performed with extracts of overexpressing plants of the splicing variants 5 and 2.9 in Hepa1c1c7-Lux cells did not confirm the gene-to-trait relationship of linking the At3g54830 overexpression with chemoprotective activity. Therefore, it should be considered that this mutant might after all not be chemoprotective. Lines overexpressing different splicing variants should be analyzed metabolically to determine the amino acid and secondary metabolite content. The identification of a functional splicing variant would enable experiments for elucidating the substrate of the transporter.

4.3.3 Mutant lines 65.32 and 67.18

The mutant lines 65.32 and 67.18 overexpressed the gene At1g63820. On the basis of *in silico* analysis the unknown protein is predicted to be soluble and it contains a CCT-domain, which possesses a putative nuclear localization signal. The nuclear localization could be experimentally confirmed via a GFP fusion construct in cell culture and tobacco.

The CCT-domain is found in 45 *A. thaliana* proteins. Among them are proteins involved in flowering (*247*), circadian rhythm regulation (*248-250*), light signaling (*251,252*) and response to sugars (*184*). It could be shown that the CCT domain of the proteins TIMING OF CAB EXPRESSION 1 and CONSTANS interacts with the transcription factor ABSCISIC ACID INSENSITIVE3 and the CCT domain of CONSTANS interacts with the proteins AtHAP2 and HAP3 to form a complex that is involved in flowering control (*253,254*). The gene At1g63820 has been categorized to the ASML2 family. This family of eight members differs from other CCT-domain protein families due to several unique characteristic residues and lack of the B-box zinc-finger domains, the pseudo receiver domain and the GATA-type zinc finger domain found in members of other CCT-domain families. The only experimentally

analyzed member of this family is the protein ASML2, which regulates a subset of sugar inducible genes (*184*).

Extracts of the At1g63820 recapitulation lines led to a strong induction of the luminescence signal in Hepa1c1c7-Lux cells. The induction was higher than that caused by extracts of line 65.32. One explanation could be differences in At1g63820 expression levels. In past screens, the extracts of 65.32 led to a stronger induction of the EpRE promotor.

A metabolite profile would allow a better understanding of the role of At1g63820 in plant secondary metabolism. Even though little information is available at this point the recapitulation of At1g63820 could confirm the chemoprotective effect of the mutant extracts. From what is known about other CCT-domain proteins, this protein may also interact with other proteins and have gene regulatory functions. It would be interesting to determine interaction partners of the protein for example via a yeast-two-hybrid screen using an *A. thaliana* transcription factor library.

4.3.4 Mutant lines 88.06, 89.29, 89.37, 90.32, 90.40, 90.43 and 90.53

The seven mutant lines 88.06, 89.29, 89.37, 90.32, 90.40, 90.43 and 90.53 overexpressed the gene At5g55880/At5g55890. Since the two genes are identical, also including parts of the intergenic regions, the insertion site could not be precisely localized. The genes At5g55880 and At5g55890 are part of a cluster of 27 genes. Within this cluster several duplication events have taken place. The mutant became of interest because of the repeated discovery of the same mutation during the screening. In addition, metabolite profiling revealed that the plants accumulated unknown compounds with long carbon chains, which were initially identified as long chain fatty acids. Further analyses did not confirm an accumulation of fatty acids. To date the unknown compounds could not be identified. Further metabolite profiling attempts will be necessary. It will be interesting to experimentally analyze the gene functions in the future, since none of the gene functions of this cluster are known. The isolation of loss-of-function mutants will be challenging, due to numerous duplication events that have take place within this cluster.

4.3.5 Mutant line *GH3.5-1D-cp*

The mutant line *GH3.5-1D-cp* overexpressed the early auxin responsive gene AtGH3.5. The rosette leaves of the mutants were curled, showing the typical morphology during auxin deficiency, which corresponds with the phenotypes of the AtGH3.5 activation tagged mutants observed by Zhang *et al.* (2008) and Park *et al.* (2007) in the *gh3.5-1D* and *wes1-D* mutants, respectively (*185,186*). An activation tagged mutant (*dfl1-D*) of the closely related AtGH3.6 with a similar phenotype has also been identified (*187,188*).

Metabolite profiling of 41.66 revealed elevated amounts of the auxin degradation product OxIAA-Glc (*255*), the aliphatic glucosinolates 7MSOH and 8MSOO, as well as the indole glucosinolate I3M. AtGH3.5 adenylates IAA and SA *in vitro* (*256*). Staswick *et al.* (2005) could demonstrate that GH3.5 conjugates the amino acids Ala, Asp, Phe, and Trp to IAA (*188*). Park *et al.* (2007) detected reduced amounts of free IAA and elevated concentrations of IAA-amino acid conjugates in *wes1-D* (*186*). In our detection system IAA amino acid conjugates were not measurable. In the *gh3.5-1D* mutant on the other hand the free IAA level as well as the IAA-amino acid conjugate levels did not differ from wild-type plants (*257*). The different IAA-amino acid conjugates influence the fate of auxin. IAA-Asp conjugates are precursors of a degradation pathway. The auxin conjugates with Ala and Phe are hydrolyzed back to free auxin, while IAA-Trp was shown to inhibit several IAA-induced growth responses (*255*). An accumulation of OxIAA-Glc was observed here. A conversion of IAA into OxIAA-Glc has been reported to be the main metabolic process of IAA in Arabidopsis (*258*). A role of AtGH3.5 in plant defense has been proposed (*185,186,257*). This is supported by the fact that in both the *wes1-D* and the *gh3.5-1D* mutants the SA content is elevated and transcript levels of PR-1 gene are enhanced (*186,257*). The activation tagged mutants of GH3.5 are more stress tolerant to both biotic and abiotic factors.

The observed altered accumulation of glucosinolates could be a product of defense signaling. Further analyses of the glucosinolate content will need to be performed in order to confirm the obtained results. The link between IAA and SA conjugation and glucosinolate accumulation needs to be clarified in future studies. From the given data it can be concluded that the enhanced amount of glucosinolates in the samples

most likely induced the EpRE reporter cells. The well-known chemoprotective effects of glucosinolates have been described already in this thesis. It has not been reported that auxin conjugates induce phase II detoxification enzymes. It cannot be ruled out that OxIAA-Glc or its further degradation products have an inducing effect on the EpRE reporter systems. To answer this question it is necessary to test the individual compounds for inducing effects in the reporter systems in the future.

4.3.6 Mutant line 68.24

The mutant line 68.24 overexpressed the gene MIR858a, which encodes the miR858a, the only member of the miR858 family. Targets of miR858a have been predicted via *in silico* analyses and partly confirmed experimentally. Addo-Quaye *et al*. (2011) published *in vitro* confirmed targets of plant microRNAs based on degradome sequencing (*190*). The authors sorted the targets into three categories, with category I being the most abundant tags matching the transcript. Using the degradome sequencing method, four targets of miR858a could be identified: MYB12 (category I), MYB111 (category II), MYB13 and MYB20 (category III). Fahlgren *et al*. (2007) identified putative targets of microRNAs using a deep sequencing and computational approach (*191*). The identified targets of miR858a are MYB12 and MYB83. MYB20 and MYB83 have been suggested to play a role in the secondary cell wall biosynthesis (*259*). The expression of MYB13 is regulated by dehydration, exogenous ABA and wounding. A role as a linker between shoot morphogenesis and environmental signaling has been suggested for MYB13 because it was detected at the shoot apex and base of developing flowers (*260-262*).

The regulatory role of the transcription factors MYB11, MYB12 and MYB111 in the phenylpropanoid pathway has been analyzed in detail. In developing seedlings of Arabidopsis it could be demonstrated that MYB12 controls the flavonol biosynthesis in the roots and MYB111 has a regulatory function mainly in the cotyledons. MYB11 has only a minor regulatory function in the flavonol biosynthesis in seedlings (*149,263*). The transcription factors activate the expression of the enzymes CHS, CHI, F3H and FLS1, which are necessary for the production of flavonols via the processing of *p*-coumaroyl-CoA to form the basic flavonol aglycone (Fig. 62).

Anthocyanins are also produced via the phenylpropanoid pathway. Their production is regulated among others by the transcription factors MYB75, MYB90, MYB113, MYB114 and MYB123 (*260*). The beneficial effect of flavonols, such as quercetin and kaempferol, has been described already in the introduction (1.1.5). These compounds are known inducers of phase II enzymes (*113,200,207*). The line 68.24 can be used to gain new insights into the regulation of the production of chemoprotective compounds in Arabidopsis.

Several plants of line 68.24 were analyzed for MYB11, MYB12 and MYB111 expression. The plant with the highest overexpression exhibited a strong downregulation of MYB11, MYB12 and MYB111, with the expression of MYB11 being the lowest, followed by MYB12 and MYB111. It could also be shown that the genes CHI, F3H, FLS, PAL1, and PAL2 were downregulated. CHI, F3H, and FLS are transcriptionally controlled by MYB11, MYB12 and MYB111 (*149,263*). Further analyses of MIR858a overexpressing lines revealed that MYB11 and MYB111 were downregulated only upon a relatively high expression of MIR858a. Taken together the data suggest that MIR858a influences the phenylpropanoid pathway by targeting the transcription factors MYB11, MYB12 and MYB111.

The expression pattern of MIR858a was analyzed using promotor:GUS lines. In developing seedlings, the expression could be detected in the vascular tissue of the cotyledons, the hypocotyl and the root meristem under short days conditions. No expression could be observed in the roots. Under high light conditions, the promotor was also active in the roots. Stracke *et al.* (2007) analyzed the promotor activity of MYB11, MYB12 and MYB111 in seedlings. The authors could find MYB11 expression in the apical meristem, the primary leaves, the apex of cotyledons, at the hypocotyl-root transition, the origin of lateral roots, root tips and vascular tissue of lateral roots. MYB12 is expressed in the root, in the vascular tissue of the hypocotyl-root transition and mildly at the region of apical meristem and cotyledon apex. MYB111 expression was found in the cotyledons and primary leaves (*149*). The parallel analysis of promMIR858a:GUS lines and MYB11, MYB12 and MYB111 promotor:GUS lines under different light conditions would help to elucidate if the genes are expressed in the same organs under given light conditions. From the given data it can be speculated that under short day conditions miR858a is expressed in the

same tissue as MYB11 and MYB111, while under more intensive light conditions MIR858a is expressed in roots and targets MYB12.

The flavonoid gene expression in Arabidopsis seedlings is controlled by a precise developmental timing mechanism. It was proposed that the link between seedling development and flavonoid synthesis helps the germinating plant to survive in the environment because the flavonoids can protect against pathogens and UV-light, especially, when the chloroplasts are not yet fully developed (*264*). The miR858a might also play a regulatory role in flavonoid compound biosynthesis during seedling development.

The promotor of MIR858a was also active in the flowers, the veins of rosette leaves and in the stems of the siliques. It could be shown that MYB11, MYB12 and MYB111 are needed for flavonol biosynthesis in the flowers but not in pollen grains. MYB111 is necessary for the production of flavonols in leaves. MYB12 and MYB111 are needed for production of flavonols in siliques (*265*). In addition, another putative target of miR858a, MYB13, is also expressed in the basis of developing flowers (*261*).

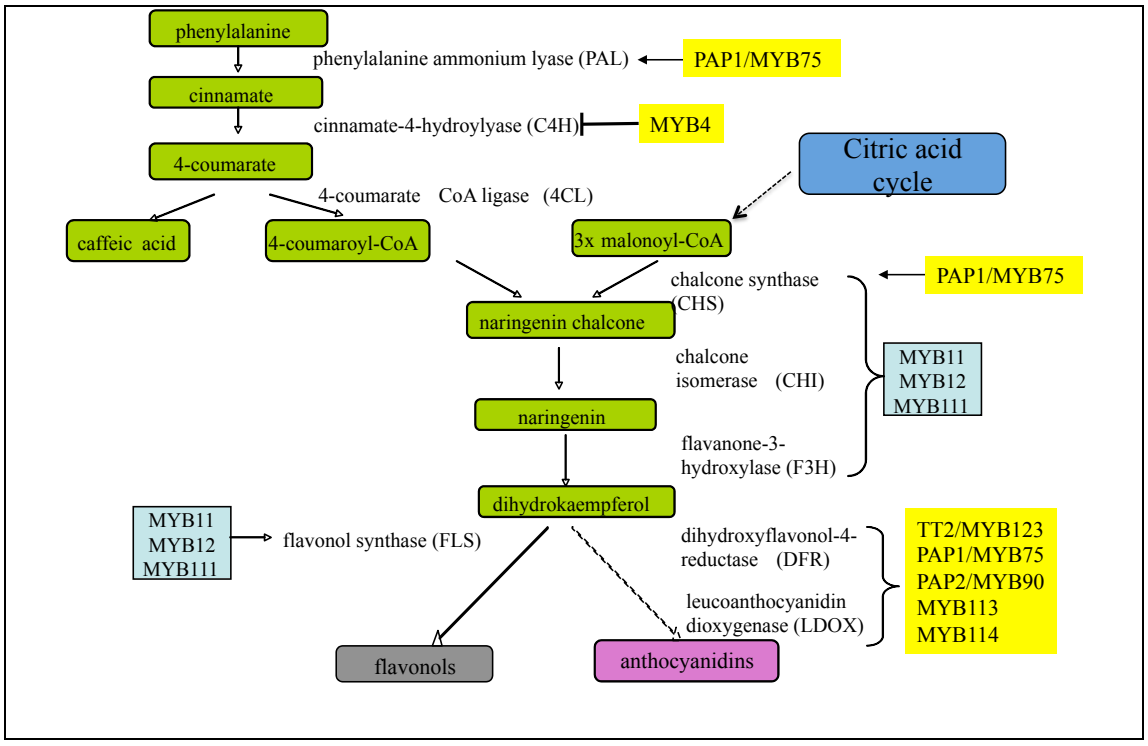


Fig. 62: A simplified schematic summary of the flavonol and anthocyanidin biosynthesis in *A. thaliana* and the role of MYB11, MYB12 and MYB111 (adapted from (*149,260*)).

The basic leucine zipper (bZIP) transcription factor ELONGATED HYPOCOTYL 5 (HY5) is involved in the regulation of light responsive genes (*266-269*). Among others, HY5 has been shown to regulate flavonoid accumulation in response to UV-B and visible light (*270,271*). The transcription factor is necessary for the activation of MYB12 and MYB111 under UV-B and visible light (*196*). Recently, it was shown by Zhang *et al.* (2011) that the promotor region of MIR858a contains light regulatory elements that are targeted by HY5 (*197,272*). A nothernblot analyses revealed reduced expression of MIR858 in a HY5 loss-of-function mutant (*hy5*) and during light-to-dark transition. These results are consistent with the observation that MIR858a was expressed differentially depending on light intensity and time of day.

The metabolite profiling of plants of line 68.24 did not yield analyzable results due to high variation between individual plants. Since the expression of MIR858a in later generations of 68.24 was low and varied between individual plants, these factors can influence the secondary metabolite accumulation in the plants.

In future analyses the miR858a accumulation itself needs to be analyzed, since the expression pattern using promotor:GUS-lines and the analysis of gene expression may not reflect actual accumulation of the mature microRNA. MicroRNAs have been shown to be able to travel from cell-to-cell and long-distance within the phloem (*273*). In addition, the tissue specific expression of miR858a should be analyzed further. As performed in another study with miR395 and its target gene (*274*), one way of analysis would be by expressing MYB11, MYB12 and MYB111 fused with GFP under the control of the native promotor and inducing MIR858a. In addition, the MIM858 lines should be subject of further analyses such as the impact of high light on the mutants. One way of analyzing the tissue specific accumulation of different flavonols in seedlings is by staining them with diphenylboric acid 2-aminoethylester. Stracke *et al.* (2007) could make a link between flavonol accumulation and the function of MYB11, MYB12 and MYB111 by using different loss-of-function mutants in their experiment (*149*). These experiments are currrently in progress with the different MIR858a mutant lines.

4.4 Conclusion and Prospects

The screening for chemoprotective activity in the TAMARA population has led to the discovery of a broad spectrum of candidate genes. 26% of the detected genes have not been assigned to a function. Many of the other detected genes could be allocated to a putative gene function only on the basis of sequence similarities but have not been characterized experimentally. Together with further metabolite profiles the detected mutants offer much potential for further discovery of genes involved in plant secondary metabolism.

In this thesis, six genes were analyzed in more detail, with emphasis on finding a link to the production of chemoprotective compounds. The β-glucosidase PYK10 is a promising enzyme for further studies both from the point of view of elucidating plant secondary metabolism and defense responses and in finding new chemoprotective compounds. The mutant *PYK10-1D-cp* enables further analyses in these directions.

A direct link between the overexpression of the putative amino acid transporter in mutant *AAT-1D-cp* and the accumulation of ITCs could not be found with the given data. The metabolite profile uncovered an accumulation of ITCs, which are strong inducers of phase II enzymes. The chemoprotective activity could not be confirmed in all screens. The overexpressing lines and the different splicing variants will be interesting for further studies.

The mutant 65.32, which overexpresses a protein of unknown function, seems to be a potent candidate for further experiments. Little is known about this CCT-domain protein family ASML2. On the basis of information known about other CCT-domain proteins and their regulatory functions, the gene At1g63820 might be an important regulator of a secondary metabolite pathway. It would be interesting to find out if the protein has transcription factor activity and to find its target genes. One way to find putative targets would be via a pull down assay. As already mentioned above the protein might also interact with other proteins.

The second gene, for which little data could be obtained, is At5g55880/At5g55890. The repeated detection of this mutant during the screening and the relatively high induction of luminescence suggest an involvement of the protein in the production or regulation of EpRE inducing compounds. Further research in this area will be worthwhile.

The well-characterized gene AtGH3.5 has been detected in two further activation tagged mutants by other groups. The enzyme conjugates amino acids to auxin and SA and plays a role in plant-pathogen interaction. The discovered glucosinolates are most likely the EpRE inducing compounds of this mutant. Nevertheless, for future analyses it will be interesting to study the fate of the OxIAA-Glc conjugates and if they exhibit chemoprotective activity. In addition, it should be considered that other compounds are affected in this mutant, such as SA, which is also conjugated by GH3.5.

The MIR858a is a candidate with a lot of potential for further studies. It is likely that phenylpropanoid derived compounds induced the EpRE reporter system. The characterization of miR858a will give further insight into the regulation of phenylpropanoid production. In addition, other MYB transcription factors should be considered as targets of miR858a.

In general, compounds with putative chemoprotective activity need to be isolated and tested individually, using the reporter cells, in order to confirm their chemoprotective activity.

5 Supplementary material

5.1 Results of the luciferase and quinone reductase assays

Table S1: Luciferase screening results of the mapped mutant lines

	luminescence /protein					
line	**mutant**	**stnd.er.**	**Col-*0***	**stnd.er.**	**ratio**	**phenotype**
39.45	74.50	±3.39	46.46	±2.55	**1.60**	like wt
	117.46	±10.27	166.08	±38.46	**0.71**	
39.46	58.00	±2.65	46.46	±2.55	**1.25**	smaller, slimmer leaves
	97.38	±11.49	149.05	±9.24	**0.65**	
40.36	91.72	±5.27	55.80	±1.13	**1.64**	like wt
	16.02	±0.47	27.69	±3.03	**0.58**	
	73.71	±3.00	37.41	±6.78	**1.97**	
	127.37	±8.82	60.85	±3.58	**2.09**	
40.76	51.95	±5.076	46.46	±2.55	**1.12**	rolled leaves
	22.29	±2.44	27.69	±3.03	**0.80**	
	23.94	±1.17	37.41	±6.78	**0.64**	
	64.34	±10.28	60.85	±3.58	**1.06**	
41.08	44.63	±2.47	44.85	±2.29	**1.00**	slightly smaller plants
41.58	86.77	±4.35	55.81	±1.13	**1.55**	curly leaves, smaller
	44.72	±1.1	37.41	±6.78	**1.20**	
	129.51	±6.66	60.85	±3.58	**2.13**	
41.66	50.09	±3.91	44.85	±2.29	**1.12**	curly leaves, smaller
	35.25	± 0.54	37.41	±6.78	**0.94**	
	142.55	±4.59	60.85	±3.58	**2.34**	
42.25	80.84	±2.59	44.85	±2.29	**1.80**	like wt
	96.93	±14.22	149.05	±9.24	**0.65**	
42.63	73.33	±5.07	50.88	±2.91	**1.44**	like wt
	202.08	±9.27	149.05	±9.24	**1.36**	
43.09	54.95	±3.2	28.71	±2.66	**1.91**	like wt
	248.77	±20.38	149.05	±9.24	**1.67**	
43.40	62.05	±1.76	51.60	±1.76	**1.20**	like wt
	42.17	±3.59	27.69	±3.03	**1.52**	
43.62	46.60	±4.03	28.71	±2.66	**1.62**	like wt
	71.78	±2.39	166.08	±38.46	**0.43**	
46.23	80.81	±0.76	125.37	±9.26	**0.64**	like wt
46.24	142.16	±10.63	48.98	±2.17	**2.90**	like wt
	120.98	±6.39	149.05	±9.24	**0.81**	
47.15	102.30	±3.85	48.98	±2.17	**2.09**	like wt
	53.65	±1.59	166.07	±38.46	**0.32**	
47.25	99.67	±3.3	104.40	±6.69	**0.95**	like wt
47.64	180.89	±25.34	104.40	±6.69	**1.73**	like wt
	103.35	±13.04	149.05	±9.24	**0.69**	
48.69	106.75	±3.04	55.80	±1.13	**1.91**	like wt
	180.24	±58.46	149.05	±9.24	**1.21**	
49.34	38.09	±1.78	46.46	±2.55	**0.82**	smaller
49.44	70.45	±5.07	55.80	±1.13	**1.26**	smaller
	313.62	±28.74	166.07	±38.46	**1.89**	
49.53	87.48	±2.83	75.05	±3.16	**1.17**	darker, earlier flowering

	luminescence	/protein				
line	**mutant**	**stnd.er.**	**Col-*0***	**stnd.er.**	**ratio**	**phenotype**
49.59	52.99	±3.58	75.05	±3.16	**0.71**	smaller
49.62	51.59	±4.34	32.62	±2.75	**1.58**	like wt
	105.20	±6.59	149.05	±9.24	**0.71**	
50.02	62.76	±4.71	104.40	±6.69	**0.60**	like wt
50.32	114.89	±9.49	104.40	±6.69	**1.10**	like wt
	67.15	±3.58	149.05	±9.24	**0.45**	
50.46	125.76	±4.99	125.37	±9.26	**1.00**	like wt
50.52	30.63	±0.75	75.05	±3.16	**0.41**	smaller
50.54	30.50	±2.73	75.05	±3.16	**0.41**	like wt
50.56	75.06	±12.16	104.39	±6.69	**0.72**	like wt
50.66	140.17	±5.977	104.40	±6.69	**1.34**	smaller
	42.51	±1.19	27.69	±3.03	**1.54**	
51.14	36.15	±1.81	75.05	±3.16	**0.48**	smaller
51.34	35.04	±3.08	75.05	±3.16	**0.47**	smaller
51.70	41.44	±0.81	75.05	±3.16	**0.55**	like wt
52.18	98.11	±4.74	104.40	±6.69	**0.94**	like wt
53.11	47.57	±4.72	75.05	±3.16	**0.63**	like wt
55.16	47.58	±0.75	75.05	±3.16	**0.63**	slightly larger
55.38	62.21	±6.37	75.04	±3.16	**0.83**	like wt
55.53	57.36	±2.79	75.05	±3.16	**0.76**	larger, darker leaves
57.31	51.20	±5.88	75.05	±3.16	**0.68**	more trichomes
61.23	101.23	±15.97	103.65	±6.36	**0.98**	
	49.63	±2.85	92.43	±6.18	**0.54**	
61.52*	50.66	±1.73	92.43	±6.18	**0.55**	
	77.57	±0.47	46.82	±1.18	**1.66**	
61.63*	39.72	±2.33	92.43	±6.18	**0.43**	smaller
	134.37	±12.42	129.71	±7.03	**1.04**	
63.01	124.63	±10.99	60.23	±2.84	**2.07**	like wt
	71.78	±3.36	30.89	±2.69	**2.32**	
63.16	157.71	±14.73	60.23	±2.84	**2.62**	like wt
	47.77	±1.19	30.89	±2.69	**1.55**	
63.17	170.46	±21.61	60.23	±2.84	**2.83**	smaller
	82.36	±4.78	33.44	±2.49	**2.46**	
63.67	48.73	±1.33	28.71	±2.66	**1.70**	like wt
	134.02	±4.58	149.05	±9.24	**0.90**	
64.15	121.78	±12.83	60.23	±2.84	**2.02**	smaller
	50.71	±0.79	33.44	±2.49	**1.52**	
64.02	48.12	±1.95	37.58	±1.8	**1.28**	like wt
	144.69	±2.93	69.83	±2.99	**2.07**	
64.22	103.24	±13.42	103.65	±6.36	**1.00**	like wt
	160.68	±27.58	60.23	±2.84	**2.67**	
64.33	174.72	±18.12	125.37	±9.26	**1.39**	like wt
	80.55	±4.41	149.05	±9.24	**0.54**	
64.47	161.34	±0.39	60.23	±2.84	**2.68**	like wt
	55.36	±0.27	30.89	±2.69	**1.79**	
64.48	183.70	±34.19	60.23	±2.84	**3.05**	like wt
	58.92	±5.82	30.89	±2.69	**1.91**	

	luminescence /protein					
line	**mutant**	**stnd.er.**	**Col-*0***	**stnd.er.**	**ratio**	**phenotype**
64.59	168.15	±25.24	60.23	±2.84	**2.79**	smaller
	57.27	±0.28	28.17	±1.72	**2.03**	
64.77	96.80	±6.44	55.80	±1.13	**1.73**	like wt
	114.86	±1.95	149.05	±9.24	**0.77**	
	24.88	±3.9	27.69	±3.03	**0.90**	
65.20	33.74	±0.1	55.80	±1.13	**0.60**	
65.28	72.04	±3.53	55.80	±1.13	**1.29**	like wt
	23.32	±1.53	27.69	±3.03	**0.84**	
65.32	70.99	±2.91	55.80	±1.13	**1.27**	like wt
	324.26	±24.03	166.08	±38.46	**1.95**	
66.71*	96.86	±4.58	92.43	±6.18	**1.05**	
	54.24	±2.76	46.82	±1.18	**1.16**	
67.18	127.64	±16.12	125.37	±9.26	**1.02**	like wt
	29.66	±1.52	27.69	±3.03	**1.07**	
67.31	127.60	±8.95	103.65	±6.36	**1.23**	like wt
	192.92	±13.67	149.05	±9.24	**1.29**	
67.51	73.64	±1.86	55.80	±1.13	**1.32**	smaller
	21.24	±1.14	27.69	±3.03	**0.77**	
67.74	183.64	±17.46	103.65	±6.36	**1.77**	like wt
	192.44	±10.58	149.05	±9.24	**1.29**	
68.15*	92.48	±1.7	31.62	±2.13	**2.92**	like wt
68.24*	32.93	±2.68	31.62	±2.13	**1.04**	like wt
	93.42	±0.79	46.82	±1.18	**2.00**	
68.41	118.97	±6.86	46.82	±1.18	**2.54**	
68.71*	22.14	±2.08	31.62	±2.13	**0.70**	
	78.79	±14.96	22.11	±1.50	**3.56**	
70.06	136.00	±12.45	55.80	±1.13	**2.44**	like wt
	158.20	±5.83	149.05	±9.24	**1.06**	
70.14	122.04	±11.37	104.40	±6.69	**1.17**	smaller
	27.59	±2.50	27.69	±3.03	**1.00**	
70.24	62.51	±4.27	55.80	±1.13	**1.12**	smaller
	24.68	±2.72	27.69	±3.03	**0.89**	
70.70	221.93	±9.79	192.75	±13.45	**1.15**	like wt
	50.29	±0.59	30.89	±2.69	**1.63**	
71.12	239.41	±7.75	121.71	±7.44	**1.97**	like wt
	69.34	±7.57	30.89	±2.69	**2.24**	
71.34	159.70	±15.98	125.06	±15.46	**1.28**	like wt
	71.08	±1.02	30.89	±2.69	**2.30**	
71.36	314.22	±22.11	121.71	±7.44	**2.58**	like wt
	47.87	±3.06	30.89	±2.69	**1.55**	
71.71	63.72	±2.03	40.30	±2.47	**1.58**	smaller
	25.08	±1.17	27.69	±3.03	**0.91**	
71.72	108.04	±5.17	40.30	±2.47	**2.68**	smaller
	23.04	±0.93	27.69	±3.03	**0.83**	
72.08	171.01	±18.09	130.33	±23.54	**1.31**	like wt
	52.65	±4.46	30.89	±2.69	**1.70**	

	luminescence	/protein				
line	**mutant**	**stnd.er.**	**Col-*0***	**stnd.er.**	**ratio**	**phenotype**
72.14	184.79	±12.07	163.90	±13.93	**1.13**	smaller
	56.90	±1.3	30.89	±2.69	**1.84**	
72.29	51.54	±0.83	40.30	±2.47	**1.28**	smaller, no flower petals
	75.25	±4.09	30.89	±2.69	**2.44**	
72.67	187.69	±11.36	163.90	±13.93	**1.15**	like wt
	86.65	±3.52	30.89	±2.69	**2.81**	
72.68	156.82	±5.75	121.71	±7.44	**1.29**	like wt
	50.02	±1.46	28.17	±1.72	**1.78**	
73.30*	16.87	±0.45	31.62	±2.13	**0.53**	
	61.13	±3.55	60.45	±5.19	**1.01**	
73.63*	20.45	±1.42	31.62	±2.13	**0.65**	
	63.73	±13.03	60.45	±5.19	**1.05**	
75.65*	28.06	±1.64	31.62	±2.13	**0.89**	smaller
	20.15	±1.00	60.45	±5.19	**0.33**	
75.66*	122.22	±2.77	22.11	±1.50	**5.53**	like wt
75.67*	149.54	±19.82	22.11	±1.50	**6.76**	
76.65*	123.59	±3.98	22.11	±1.50	**5.59**	
79.23*	32.02	±2.76	31.62	±2.13	**1.01**	like wt
	53.32	±5.95	23.12	±0.55	**2.31**	
81.21*	26.12	±5.13	28.95	±6.15	**0.90**	smaller
	16.74	±1.41	23.12	±0.55	**0.72**	
81.77*	40.29	±3.03	28.95	±6.15	**1.39**	like wt
	108.70	±6.71	23.12	±0.55	**4.70**	
82.20*	27.40	±1.98	23.12	±0.55	**1.19**	like wt
	20.70	±4.01	28.95	±6.15	**0.72**	
84.20*	29.19	±1.45	27.83	±2.73	**1.05**	
	30.91	±3.21	19.11	±1.74	**1.62**	
84.27*	76.90	±8.37	27.83	±2.73	**2.76**	like wt
	152.31	±8.12	19.11	±1.74	**7.97**	
84.63*	51.41	±3.76	27.83	±2.73	**1.85**	like wt
	64.35	±3.33	39.29	±0.98	**1.64**	
84.64*	15.28	±0.65	34.04	±1.89	**0.45**	like wt
	38.93	±5.43	39.29	±0.98	**0.99**	
84.75*	13.16	±0.20	34.67	±0.89	**0.38**	smaller
	87.52	±6.20	21.41	±0.6	**4.09**	
86.25*	285.16	±13.62	33.68	±2.87	**8.47**	like wt
	77.56	±4.22	39.29	±0.98	**1.97**	like wt
86.56*	119.48	±14.12	39.29	±0.98	**3.04**	
86.71*	71.53	±5.11	39.29	±0.98	**1.82**	larger, pointy leaves
86.76*	120.68	±1.58	40.48	±1.49	**2.98**	like wt
	108.05	±7.50	39.29	±0.98	**2.75**	
87.09*	83.84	±6.51	39.29	±0.98	**2.13**	like wt
87.17*	61.54	±3.13	41.48	±1.8	**1.48**	smaller
87.42*	155.68	±11.84	41.48	±1.8	**3.75**	smaller
87.58*	123.50	±6.73	41.48	±1.8	**2.98**	like wt
87.59*	88.31	±5.69	39.37	±0.83	**2.24**	like wt
88.05*	56.18	±8.63	39.37	±0.83	**1.43**	smaller

	luminescence	/protein				
line	**mutant**	**stnd.er.**	**Col-*0***	**stnd.er.**	**ratio**	**phenotype**
88.06*	88.02	±8.67	39.37	±0.83	**2.24**	smaller
88.36*	158.29	±26.15	39.37	±0.83	**4.02**	like wt
88.75*	128.08	±7.11	39.37	±0.83	**3.25**	smaller
89.01*	113.81	±6.19	39.68	±2.71	**2.87**	smaller
89.29*	128.22	±24.46	39.68	±2.71	**3.23**	like wt
89.37*	61.71	±15.29	39.68	±2.71	**1.56**	smaller
89.68*	110.54	±15.64	39.68	±2.71	**2.79**	smaller
90.15*	68.57	±3.92	39.68	±2.71	**1.73**	
90.32*	145.86	±1.46	47.85	±5.23	**3.05**	smaller
90.36*	127.22	±15.64	47.85	±5.23	**2.66**	smaller
90.37*	127.63	±20.35	47.85	±5.23	**2.67**	smaller
90.40*	211.98	±23.98	47.85	±5.23	**4.43**	like wt
90.43*	170.40	±18.58	47.85	±5.23	**3.56**	like wt
90.46*	103.73	±27.18	47.85	±5.23	**2.17**	like wt
90.53*	96.16	±15.01	45.47	±4.79	**2.12**	smaller
90.54*	144.55	±11.8	45.47	±4.79	**3.18**	like wt
90.63*	140.34	±21.83	45.47	±4.79	**3.09**	like wt
90.64*	186.71	±39.88	45.47	±4.79	**4.11**	like wt
	194.12	±30.44	45.47	±4.79	**4.27**	
	115.37	±10.03	45.47	±4.79	**2.54**	
94.12 *	74.68	±7.20	129.71	±7.03	**0.58**	dwarfish
	111.27	±8.45	129.71	±7.03	**0.86**	
	57.84	±5.10	129.71	±7.03	**0.45**	
94.31 *	82.49	±2.68	129.71	±7.03	**0.64**	very small
94.32	61.07	±6.81	50.36	±3.71	**1.21**	
94.48*	58.38	±8.00	129.71	±7.03	**0.45**	very small
94.49 *	62.57	±1.03	129.71	±7.03	**0.48**	
94.52 *	58.87	±6.97	77.83	±3.60	**0.76**	dwarfish
94.67 *	58.25	±12.54	77.83	±3.6	**0.75**	
95.16	85.38	±7.90	43.08	±1.95	**1.98**	
95.18	65.38	±8.44	43.08	±1.95	**1.52**	
95.19	74.62	±3.99	43.08	±1.95	**1.73**	
95.22	89.23	±4.33	43.08	±1.95	**2.07**	
95.31	81.54	±6.22	43.08	±1.95	**1.89**	
95.36	57.89	±6.10	16.32	±0.61	**3.55**	

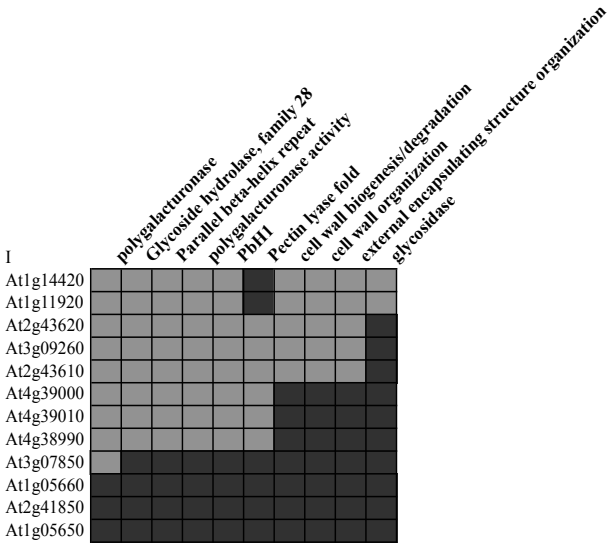
Note: Part of the screening was performed by Eigen, C. and published in his diploma thesis (2009). The data from these screenings are marked with an asterisk (*).

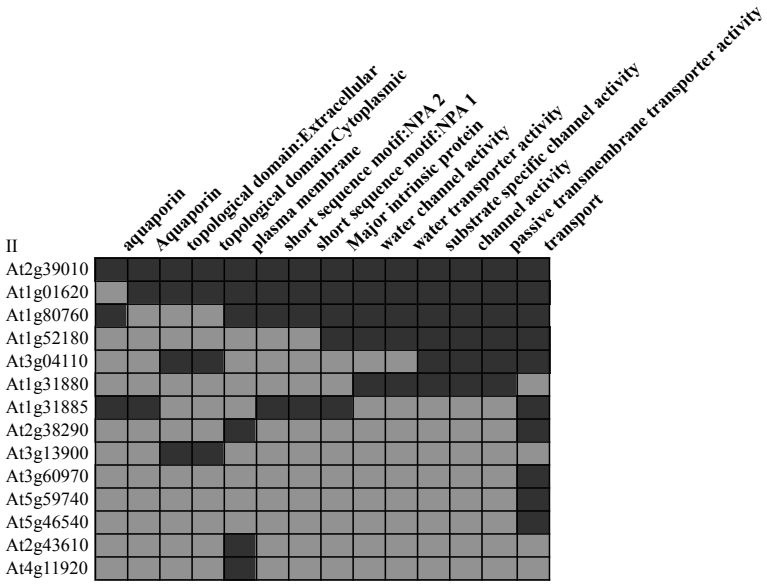
Table S2: Results of the QR-assay

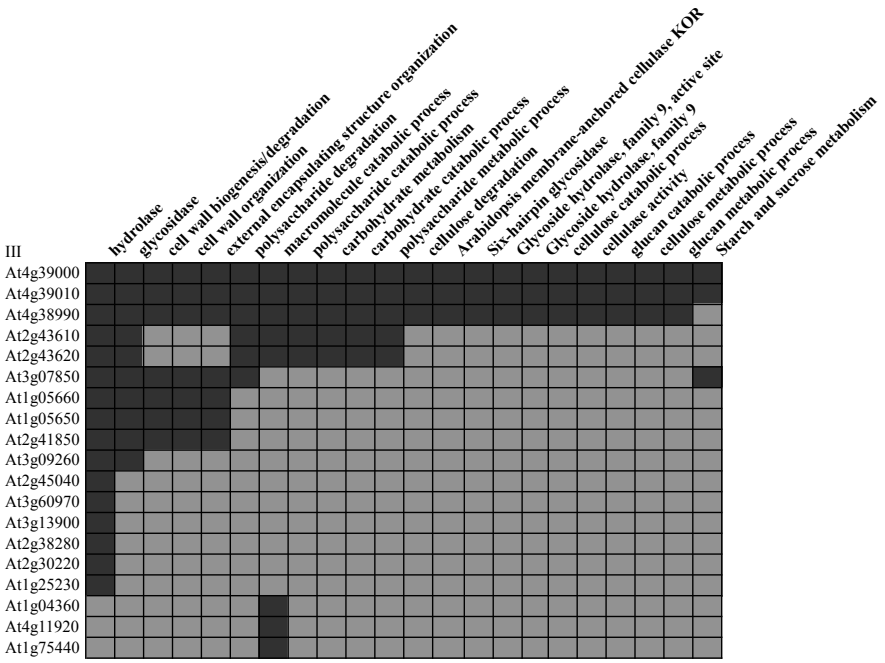
QR	mutant	stnd.er.	Col-*0*	stnd.er.	ratio
39.45	0.35	±0.005	0.27	±0.009	1.31
39.46	0.37	±0.009	0.27	±0.009	1.37
40.36	0.52	±0.009	0.39	±0.016	1.32
40.76	0.33	±0.008	0.27	±0.009	1.23
42.25	0.40	±0.008	0.27	±0.009	1.50
42.63	0.37	±0.003	0.27	±0.009	1.40
43.09	0.39	±0.009	0.40	±0.001	0.96
43.40	0.32	±0.006	0.26	±0.004	1.25
43.62	0.41	±0.012	0.40	±0.001	1.02
46.23	0.31	±0.003	0.31	±0.018	1.00
46.24	0.20	±0.030	0.27	±0.009	0.74
47.15	0.39	±0.004	0.27	±0.009	1.47
47.64	0.20	±0.003	0.22	±0.012	0.94
48.69	0.68	±0.039	0.39	±0.016	1.73
49.44	0.38	±0.006	0.27	±0.009	1.41
49.53	0.36	±0.005	0.33	±0.007	1.10
49.62	0.34	±0.045	0.36	±0.007	0.94
50.02	0.26	±0.008	0.22	±0.012	1.21
50.32	0.25	±0.008	0.22	±0.012	1.15
50.46	0.31	±0.013	0.31	±0.018	1.01
50.52	0.26	±0.006	0.33	±0.007	0.79
50.56	0.23	±0.012	0.22	±0.012	1.05
50.66	0.17	±0.008	0.22	±0.012	0.76
51.34	0.28	±0.007	0.33	±0.007	0.84
55.38	0.34	±0.006	0.33	±0.007	1.02
61.23	0.21	±0.008	0.29	±0.025	0.71
63.01	0.30	±0.003	0.31	±0.005	0.95
63.67	0.44	±0.022	0.40	±0.001	1.09
64.22	0.21	±0.014	0.29	±0.024	0.71
64.33	0.24	±0.011	0.31	±0.018	0.77
64.77	0.29	±0.012	0.31	±0.005	0.92
65.20	0.25	±0.005	0.30	±0.003	0.86
65.28	0.28	±0.003	0.31	±0.005	0.90
65.32	0.29	±0.006	0.30	±0.003	0.97
67.18	0.23	±0.013	0.31	±0.018	0.76
67.31	0.29	±0.009	0.29	±0.025	1.00
67.51	0.32	±0.004	0.31	±0.005	1.02
67.74	0.26	±0.009	0.29	±0.024	0.88
70.06	0.32	±0.005	0.30	±0.003	1.10
70.14	0.21	±0.004	0.29	±0.012	0.73

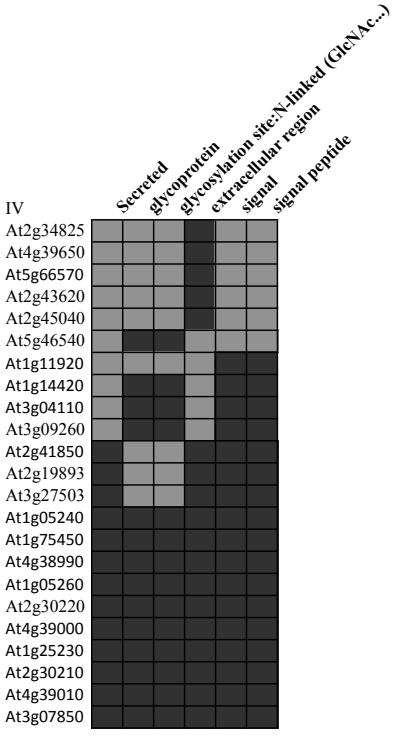
QR	mutant	stnd.er.	Col-*0*	stnd.er.	ratio
70.24	0.30	±0.004	0.30	±0.003	1.01
70.70	0.41	±0.016	0.44	±0.014	0.93
71.71	0.27	±0.022	0.36	±0.005	0.75
71.72	0.32	±0.009	0.36	±0.005	0.90
72.08	0.45	±0.038	0.44	±0.035	1.02
72.14	0.39	±0.015	0.41	±0.016	0.96
72.29	0.29	±0.009	0.36	±0.005	0.81
72.67	0.42	±0.014	0.41	±0.016	1.02

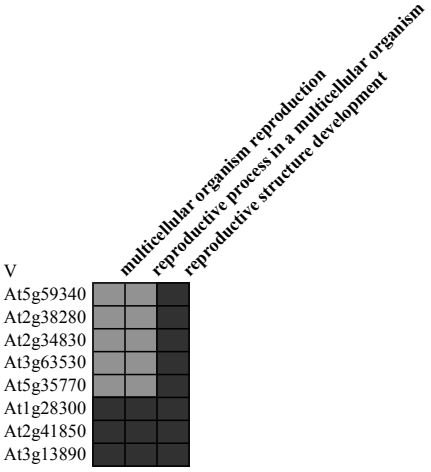
5.2 Results of the functional annotation clustering

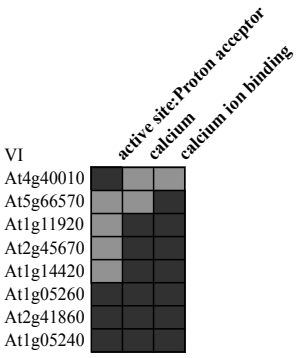


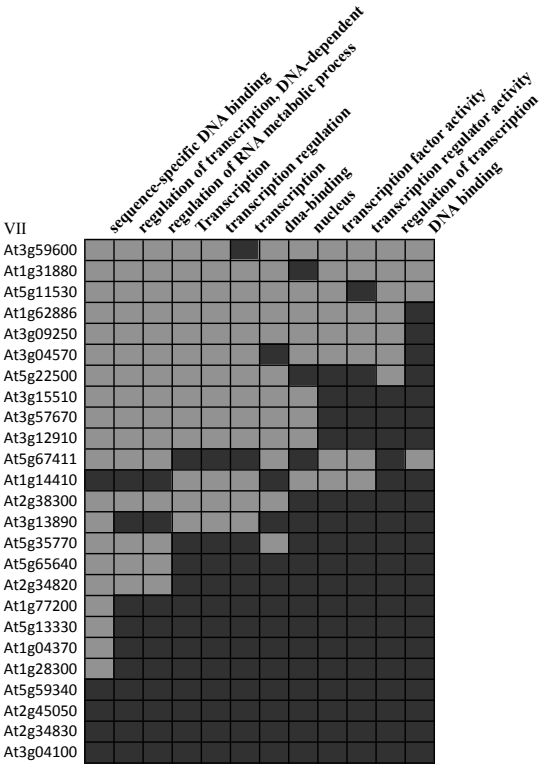


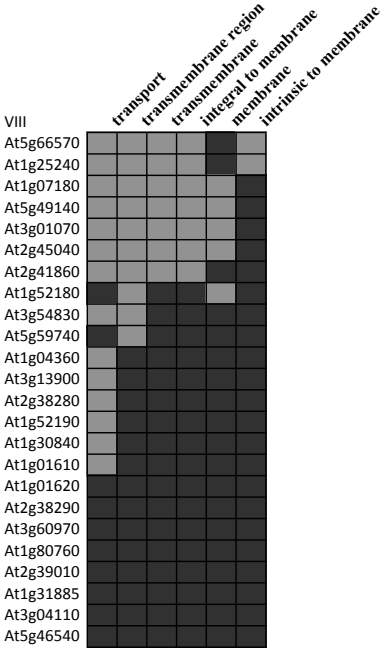

III
hydrolase
glycosidase
cell wall biogenesis/degradation
cell wall organization
external encapsulating structure organization
polysaccharide degradation
macromolecule catabolic process
polysaccharide catabolic process
carbohydrate metabolism
carbohydrate catabolic process
polysaccharide metabolic process
cellulose degradation
Arabidopsis membrane-anchored cellulase KOR
Six-hairpin glycosidase
Glycoside hydrolase, family 9, active site
Glycoside hydrolase, family 9
cellulose catabolic process
cellulase activity
glucan catabolic process
cellulose metabolic process
glucan metabolic process
Starch and sucrose metabolism
At4g39000
At4g39010
At4g38990
At2g43610
At2g43620
At3g07850
At1g05660
At1g05650
At2g41850
At3g09260
At2g45040
At3g60970
At3g13900
At2g38280
At2g30220
At1g25230
At1g04360
At4g11920
At1g75440

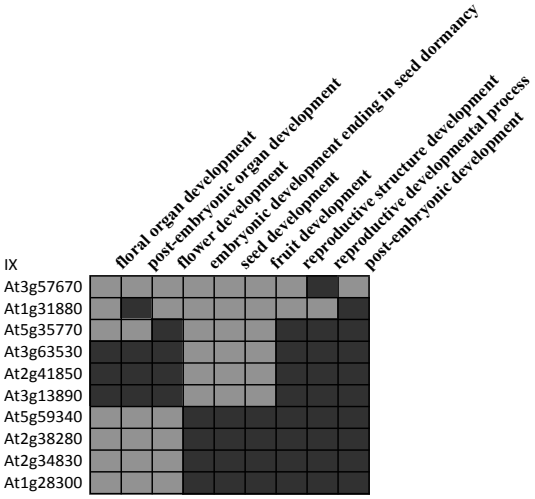

IV
Secreted
glycoprotein
glycosylation site:N-linked (GlcNAc...)
extracellular region
signal
signal peptide
At2g34825
At4g39650
At5g66570
At2g43620
At2g45040
At5g46540
At1g11920
At1g14420
At3g04110
At3g09260
At2g41850
At2g19893
At3g27503
At1g05240
At1g75450
At4g38990
At1g05260
At2g30220
At4g39000
At1g25230
At2g30210
At4g39010
At3g07850

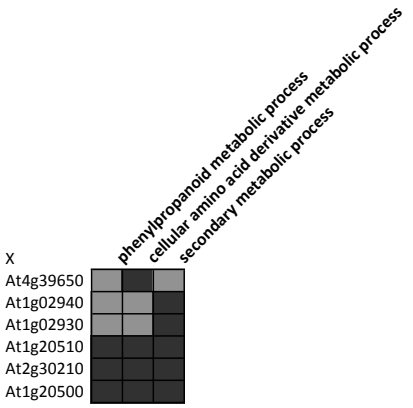

V
multicellular organism reproduction
reproductive process in a multicellular organism
reproductive structure development
At5g59340
At2g38280
At2g34830
At3g63530
At5g35770
At1g28300
At2g41850
At3g13890

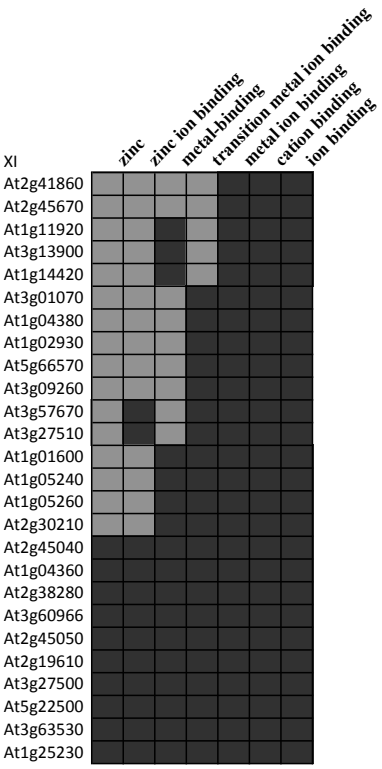

VI
active site:Proton acceptor
calcium
calcium ion binding
At4g40010
At5g66570
At1g11920
At2g45670
At1g14420
At1g05260
At2g41860
At1g05240

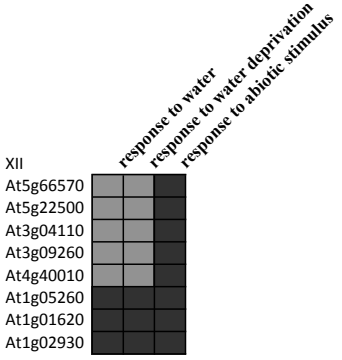

VII
sequence-specific DNA binding
regulation of transcription, DNA-dependent
regulation of RNA metabolic process
Transcription
transcription regulation
transcription
dna-binding
nucleus
transcription factor activity
transcription regulator activity
regulation of transcription
DNA binding
At3g59600
At1g31880
At5g11530
At1g62886
At3g09250
At3g04570
At5g22500
At3g15510
At3g57670
At3g12910
At5g67411
At1g14410
At2g38300
At3g13890
At5g35770
At5g65640
At2g34820
At1g77200
At5g13330
At1g04370
At1g28300
At5g59340
At2g45050
At2g34830
At3g04100

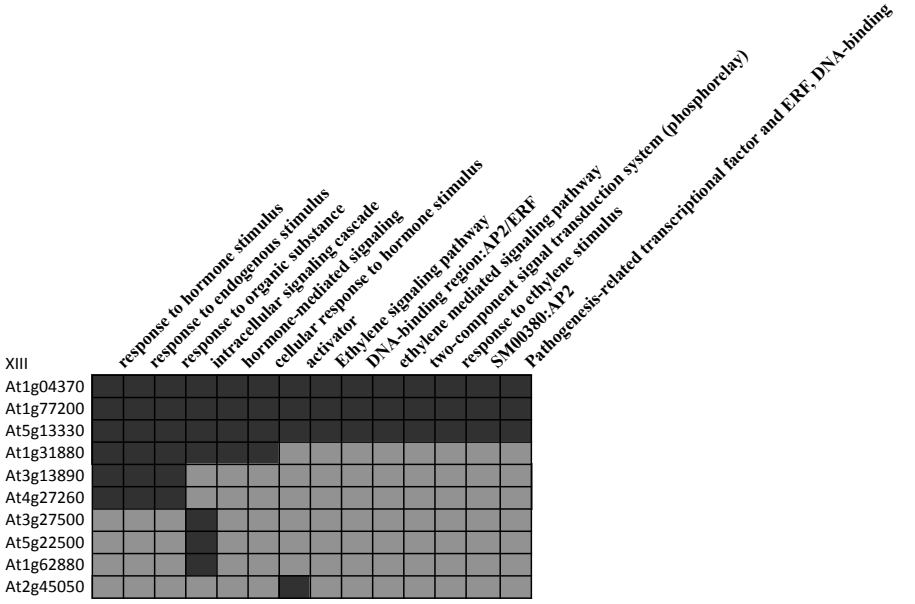

VIII
transport
transmembrane region
transmembrane
integral to membrane
membrane
intrinsic to membrane
At5g66570
At1g25240
At1g07180
At5g49140
At3g01070
At2g45040
At2g41860
At1g52180
At3g54830
At5g59740
At1g04360
At3g13900
At2g38280
At1g52190
At1g30840
At1g01610
At1g01620
At2g38290
At3g60970
At1g80760
At2g39010
At1g31885
At3g04110
At5g46540

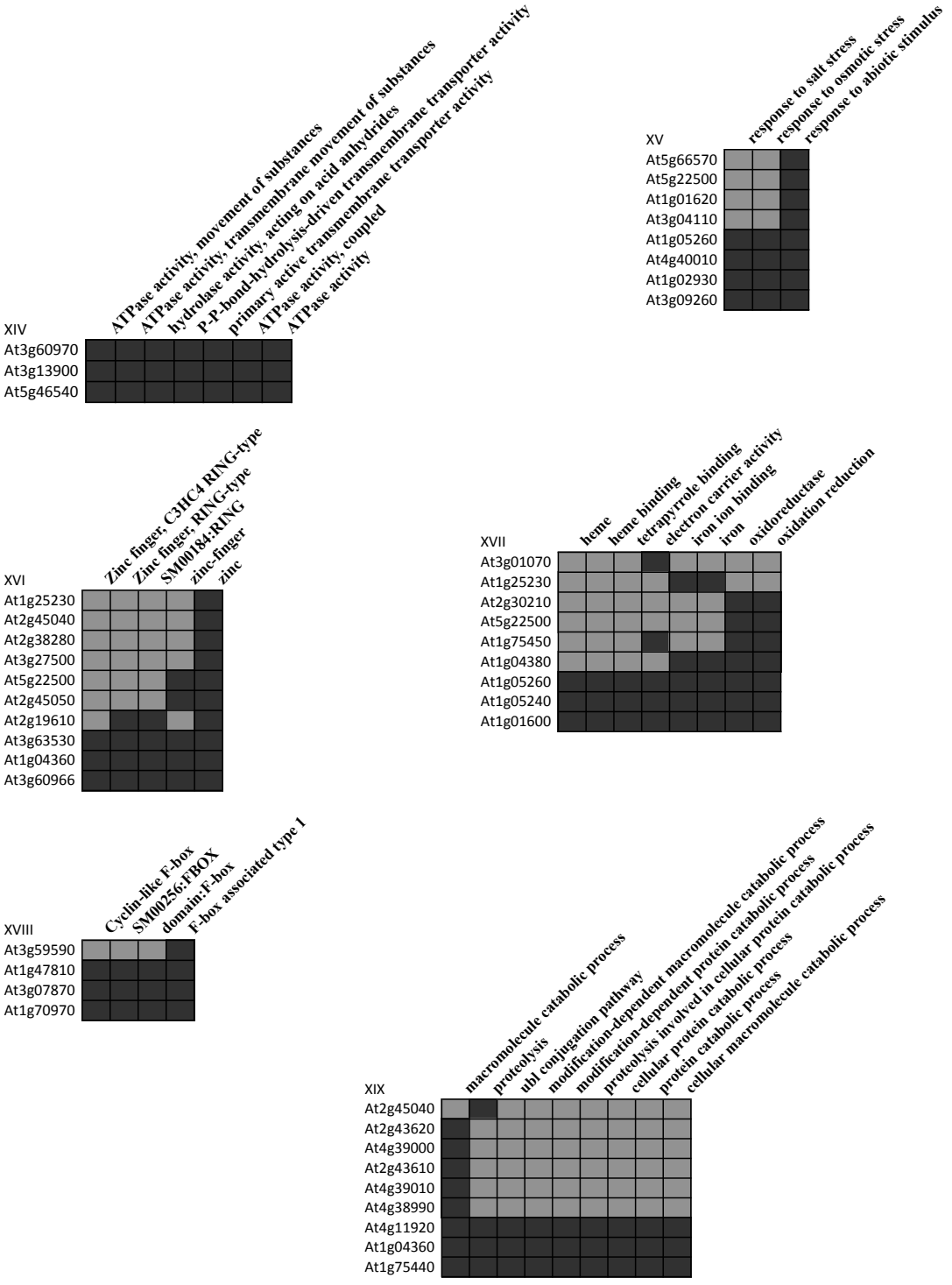

IX
floral organ development
post-embryonic organ development
flower development
embryonic development ending in seed dormancy
seed development
fruit development
reproductive structure development
reproductive developmental process
post-embryonic development
At3g57670
At1g31880
At5g35770
At3g63530
At2g41850
At3g13890
At5g59340
At2g38280
At2g34830
At1g28300


phenylpropanoid metabolic process
cellular amino acid derivative metabolic process
secondary metabolic process
X
At4g39650
At1g02940
At1g02930
At1g20510
At2g30210
At1g20500


zinc
zinc ion binding
metal-binding
transition metal ion binding
metal ion binding
cation binding
ion binding
XI
At2g41860
At2g45670
At1g11920
At3g13900
At1g14420
At3g01070
At1g04380
At1g02930
At5g66570
At3g09260
At3g57670
At3g27510
At1g01600
At1g05240
At1g05260
At2g30210
At2g45040
At1g04360
At2g38280
At3g60966
At2g45050
At2g19610
At3g27500
At5g22500
At3g63530
At1g25230


response to water
response to water deprivation
response to abiotic stimulus
XII
At5g66570
At5g22500
At3g04110
At3g09260
At4g40010
At1g05260
At1g01620
At1g02930


response to hormone stimulus
response to endogenous stimulus
response to organic substance
intracellular signaling cascade
hormone-mediated signaling
cellular response to hormone stimulus
activator
Ethylene signaling pathway
DNA-binding region:AP2/ERF
ethylene mediated signaling pathway
two-component signal transduction system (phosphorelay)
response to ethylene stimulus
SM00380:AP2
Pathogenesis-related transcriptional factor and ERF, DNA-binding
XIII
At1g04370
At1g77200
At5g13330
At1g31880
At3g13890
At4g27260
At3g27500
At5g22500
At1g62880
At2g45050


ATPase activity, movement of substances
ATPase activity, transmembrane movement of substances
hydrolase activity, acting on acid anhydrides
P-P-bond-hydrolysis-driven transmembrane transporter activity
primary active transmembrane transporter activity
ATPase activity, coupled
ATPase activity
XIV
At3g60970
At3g13900
At5g46540
response to salt stress
response to osmotic stress
response to abiotic stimulus
XV
At5g66570
At5g22500
At1g01620
At3g04110
At1g05260
At4g40010
At1g02930
At3g09260
Zinc finger, C3HC4 RING-type
Zinc finger, RING-type
SM00184:RING
zinc-finger
zinc
XVI
At1g25230
At2g45040
At2g38280
At3g27500
At5g22500
At2g45050
At2g19610
At3g63530
At1g04360
At3g60966
heme
heme binding
tetrapyrrole binding
electron carrier activity
iron ion binding
iron
oxidoreductase
oxidation reduction
XVII
At3g01070
At1g25230
At2g30210
At5g22500
At1g75450
At1g04380
At1g05260
At1g05240
At1g01600
Cyclin-like F-box
SM00256:FBOX
domain:F-box
F-box associated type 1
XVIII
At3g59590
At1g47810
At3g07870
At1g70970
macromolecule catabolic process
proteolysis
ubl conjugation pathway
modification-dependent macromolecule catabolic process
modification-dependent protein catabolic process
proteolysis involved in cellular protein catabolic process
cellular protein catabolic process
protein catabolic process
cellular macromolecule catabolic process
XIX
At2g45040
At2g43620
At4g39000
At2g43610
At4g39010
At4g38990
At4g11920
At1g04360
At1g75440

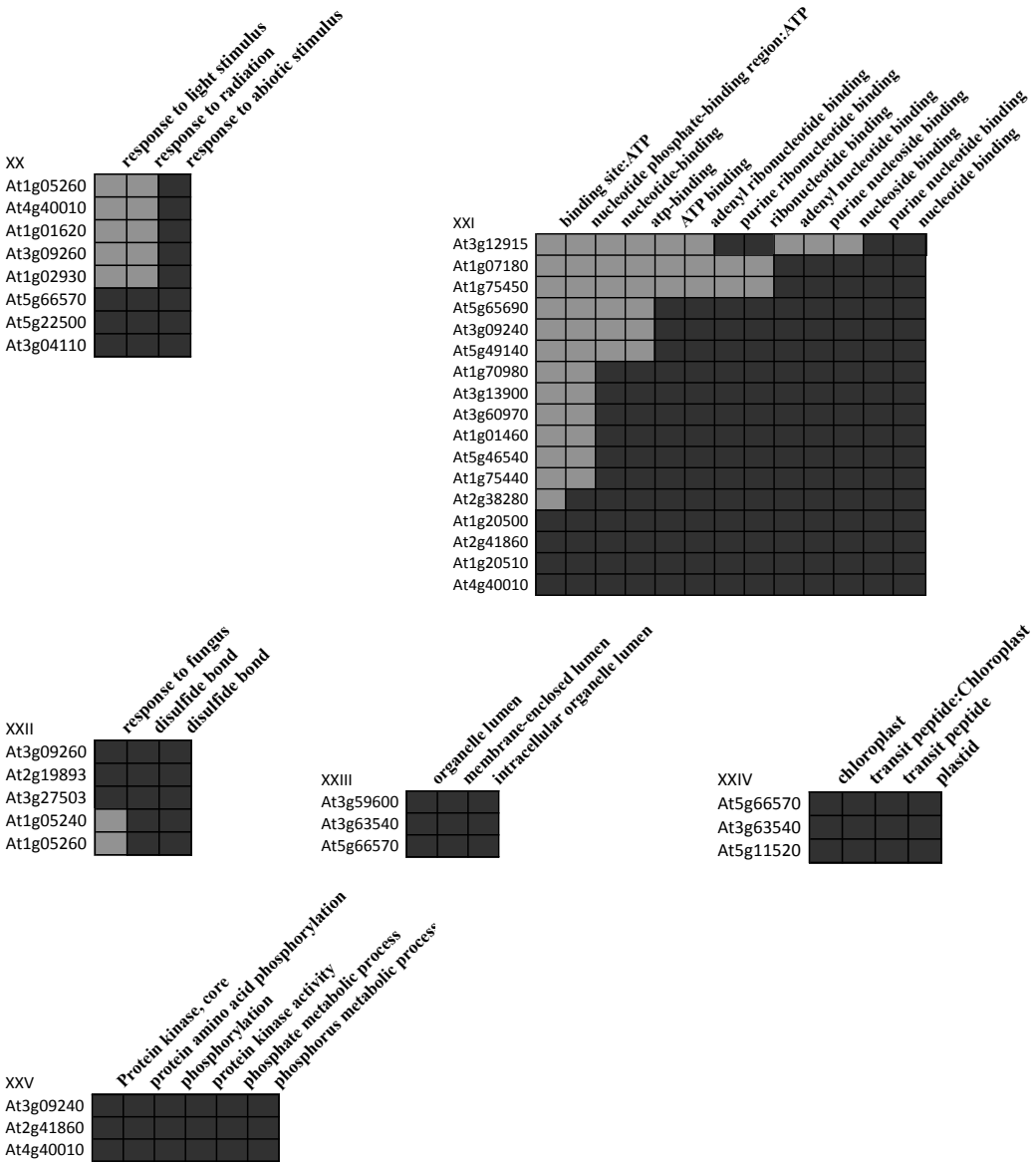


Fig. S1: Diagrams of the 25 gene clusters formed after Functional annotation clustering using DAVID. If a gene was positively reported with the specific term, the box is marked in dark grey. Light grey boxes show that the respective gene was not annotated with this term.

5.3 Subcellular localization studies of further genes

Mutant 47.64
T-DNA insertion: chromosome 4 at position 17551869
Genes cloned: At4g37290 and At4g37295
Subcellular localization: vacuole

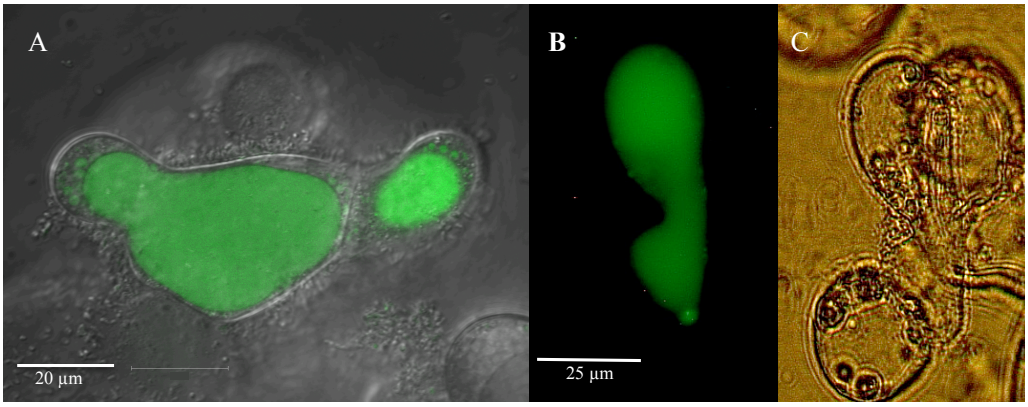


Fig. S2: Confocal microscopy of *A. thaliana* cultured cells transformed with (A) 35S: At4g37290-GFP and (B, C) 35S: At4g37295-GFP.

Mutant 57.31
T-DNA insertion: chromosome 5 at position 23932995
Gene cloned: At5g59330
Subcellular localization: cytoplasm, around the nucleus, aggregates

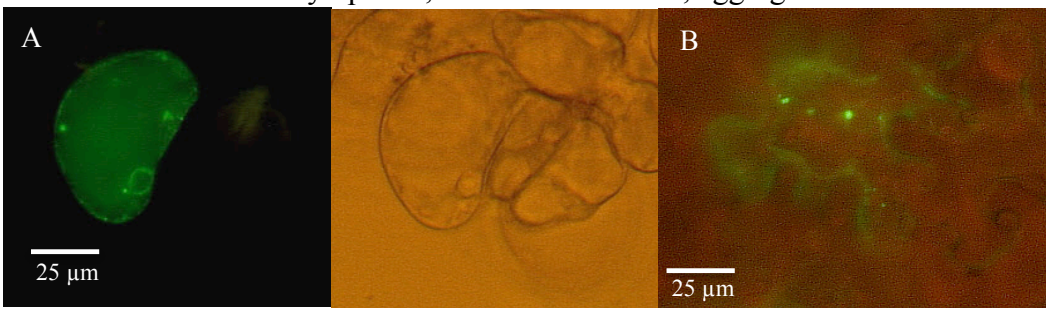


Fig. S3: Wide-field fluorescence microscopy of 35S:At5g59330-GFP transiently expressed in (A) *A. thaliana* cultured cells and (B) *N. benthamiana* leaves.

Mutant 70.70
T-DNA insertion: chromosome 5 at position 19916110, first exon of At5g49130
Gene cloned: At5g49120
Subcellular localization: nucleus

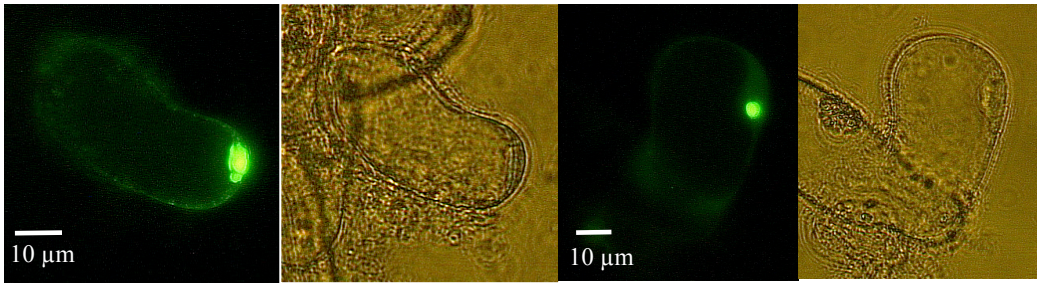


Fig. S4: Wide-field fluorescence microscopy of *A. thaliana* cultured cells transiently transformed with 35S: At5g49120-GFP

Mutant 63.16

T-DNA insertion: chromosome 1 between the genes At1g08670 and At1g08680

Gene cloned: At1g08670

Subcellular localization: nucleus, endomembrane system

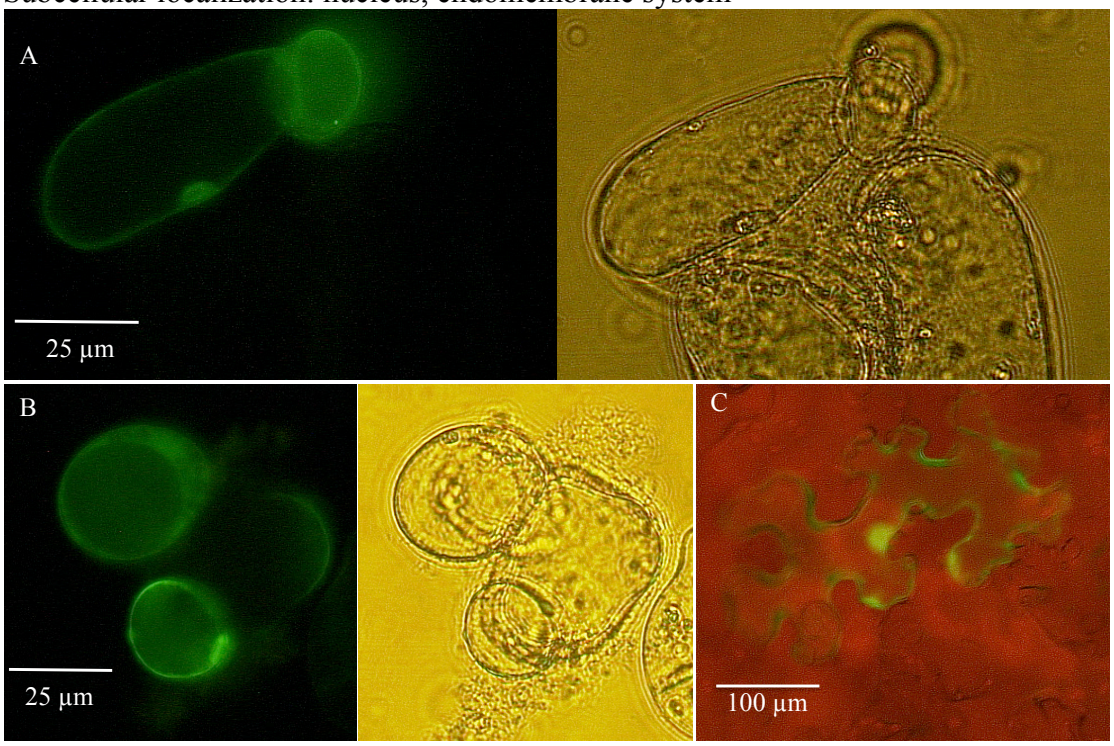


Fig. S5: Wide-field fluorescence microscopy of *A. thaliana* cultured cells and *N. benthamiana* transiently transformed with 35S: At1g08670-GFP

Mutant 51.70

T-DNA insertion: chromosome 1 between genes At1g01470 and At1g01460

Gene cloned: At1g01470

Subcellular localization: around the nucleus and in cytosol

Fig. S6: Wide-field fluorescence microscopy of (A) *N. benthamiana* and (B) *A. thaliana* cultured cells transiently transformed with 35S: At1g01470-GFP.

Table S3: List of oligonucleotides used for Cloning

Name	Sequence (5'-3')
At5g55880EcoFor	AAAGAATTCAATGTCTCTGCTTCTCAACCA
A5g55880NotRev	TTTGCGGCCGCTTGTTGTCTATATTTTGAG
At3g09260EcoRIFor	AAAGAATTCAATGGTTTTGCAAAAGCTTCCTCTCATTGG
At3g09260NotIRev	TTTGCGGCCGCTTAAGCTCATCCTTCTTGAGC
At1g08670forBamHI	AAAGGATCCAAATGTCTAAGCTTATGCAGAATCCGGTGTT
At1g08670revNotI	TTTGCGGCCGCTTAGTCGAAGAAGAAAGCAAGAGTTGAG
At1g01470BamHIFor	AAAGGATCCAAATGGCGAGCTTGCTAGATAAAGCC
At1g01470NotIrev	TTTGCGGCCGCTTGAAGAAATCTTTAAAAGTAGGAAG
At5g59330NcoIfor	AAACCATGGCTTTGGCTCTCAGG
At5g59330NotIrev	TTTGCGGCCGCTTGGAAGACCAGAGGCTAGACTTGG
At5g49120NcoIfor	AAACCATGGTGGGACTAAGTATTGTTTTG
At5g49120NotIrev	TTTGCGGCCGCTTGTAAGCGAAACCGCCTGCTTGG
At3g54830 attB1	GGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAATCACTCCACTTCTGACC
At3g54830 attB2	GGGACCACTTTGTACAAGAAAGCTGGGTCATGATCTTTGCAAGGGCTGAAT
PromshAt3g54830attB2	GGGACCACTTTGTACAAGAAAGCTGGGTCGTCGAGGCAGTACCGCAGAA
PromshAt3g54830attB1	GGGACAAGTTTGTACAAAAAAGCAGGCTTCTGAACGAAGCAACAGTCGACAC
At4g37290ecorIfor	AAAGAATTCATGATGATGAACAAAA
At4g37290ecorIrev	TTTGAATTCTTGTGGCCCGGTCCG
At4g37295ecorIfor	AAAGAATTCATGAGACCGGTTGGCT
At4g37295ecorIrev	TTTGAATTCTTGTGACCAGCACCTC
At1g63820forattB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGATTCTGATCTTTATATCT
At1g63820revattB2	GGGGACCACTTTGTACAAGAAAGCTGGGTCCCAGTAGGAATTTGTCATGT
Promshort At1g71002 BamHIfor	AAGGATCCAATCATTGTACTATCAGTAACGGCACAAC
Promshort At1g71002 NotIrev	TTTGCGGCCGCCTCAGAAATCAAAAGTTTCGATCGTCTA
Prommid At1g71002 BamHIfor	AAAGGATCCTAATTCTTTGCTCTATCGAATGACCTAA
Attb1_At1g71002fw	GGGGACAAGTTTGTACAAAAAGCAGGCTTCCTCATCTTTGTTTCGTTGTCTGT
Attb2_At1g71002 rv	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTCAGAAATCAAAAGTTTCGATCGT

Table S4: List of oligonucleotides used for mapping

name	sequence (5'-3')
LR32	ACTCGATTCTCAACCCGAAAGTATAGATCCCA
LR26	ACTCGATTCTCAAACCGAAAGTATAG
APL16	P-TATGGGATCACATTAA-amino (NH_2)
APL17	P-CGTGGGATCACATTAA-amino (NH_2)
En230as	AGAAGCACGACGGCTGTAGAATAGGA
En82as	ACCCACACTTTTACTTCGATTAAGAGTGT
En8130s	GAGCGTCGGTCCCCACACTTCTATAC
En8153s	TACGAATAAGAGCGTCCATTTTAGAGTG

Table S5: List of oligonucleotides used for confirmation of the T-DNA insertion site

mutant	primer	sequence
41.66	41.66/Hin6 Fw	AAGAGCTTTTGCCCTATTGTAGTGG
41.66	41.66/Hin6 Rv	TCAATAAATAACACACATGGGAGGC
47.64	6433F	TGAGTCGGATCTCATGTGTTT
47.64	6433R	TACGGTCGTGGATCATCGACT
50.46	g_At2g43620 FW	TTTCTGATTGGTGTCTCTTAGGAAT
50.46	g_At2g43620 RV	TCCGATTTGAGTTAAAGTATTGCG
51.14	51.14/Csp Fw	CAAGCTGCAAAACAAAGAACCGT
51.14	51.14/Csp Rv	CCAATCGCCAACCTCAGAAATA
52.18	52.18 Hin6/53.56 Csp Fw	TCTCATTACGCCAAACCTTCTCC
52.18	52.18 Hin6/53.56 Csp Rv	GCAAAAACAAACACCGATCACTAAA
59.67	g_At1g05240 FW	ATTGTTCGTGGTGTCACAGTACAAT
59.67	g_At1g05240 RV	GACGAGAGCGCAACTAGAGATTC
63.16	6316F	TGGATGAGTGATCGCTGCCC
63.16	6316R	TTATCTAGGTTTTCAATCAATCGTTTCATT
64.22	6422F	CAAGAAATGCAGCACAAGGCT
64.22	6422R	ATGAGACTTGTGCTTGTAGCGG
64.22	g_At2g19893 FW	ATCGTTCATCTGGCAACTAATATCT
64.22	g_At2g19893 RV	GATCGAGGAATCCTTATCAAACA
64.48	g_At1g70970 FW	GCATGGAAGCAAATATTAGCCAG
64.48	g_At1g70970 RV	TGTAATTGTATGGGCTTATTCATGG
65.32	6532F	AATCGAACAGGGTATAGTTTAAGCTGT
65.32	6532R	ATTCACCGAAAACAAGGATAATCCAC
66.71	g_At3g09250 FW	AGAAAGAATCCGGCAAGTATTACAA
66.71	g_At3g09250 RV	CGATGCAATCACTATGGTCAAAAT
68.24	g_At1g71002 FW	TCAAGTTTTGTCGTAAGCCTTCT
68.24	g_At1g71002 RV	ATTTGCTCGAACATGAAGACTAAA
70.06	706F	TATTCGCATACGGAGAGCAAAA
70.06	706R	TCCGGTGGATTGTTCTAGCG
70.70	7070F	ACACTAACAATGGTGGTCGAAGAA
70.70	7070R	TGTAGCAAAGGAGAAGCGACAA
72.14	7214F	CGAGCCATGTACCGGAAACTT
72.14	7214R	CAGTCAACTTAAGACCCCCGTC
72.29	7229F	TCCAGACGACGAAGTACGTGAC
72.29	7229R	AACCCAATAAAAAACACGGCC
81.77	g_At2g37450 FW	ACACTGAAACACCAGCCATGAGT
81.77	g_At2g37450 RV	CCTCGTTTTTTCACATATACAAGCT
87.58	g_At5g10180 FW	GGGATTACTGCTTTGCTGGAACT
87.58	g_At5g10180 RV	GTCCTTCCACTAATTTGATTCACC
89.29	g_At5g55880 FW	TATTATTAGTGGAGGGTTTGACCG
89.29	g_At5g55880 RV	ACGAGAGGAGTAGAAACAAGGGTT
90.46	g_At5g09530 FW	CCGTCCAAAAGATGACAAA
90.46	g_At5g09530 RV	ATGGAGACTGAAAAGCCTG

Table S6: List of oligonucleotides used for mutant screening with real time PCR

mutant	name	sequence (5'-3')
39.45	At1g80760F	TTCTTGGAGCGTTAATCGGA
39.45	At1g80760R	TTTGGGTGCTTCATCTTCCTCT
39.45	At1g80770F	ACAGAACAAGGTCTCAATGAGCTAAA
39.45	At1g80770R	TGGATCTTCTCCATCTCACTACTCAG
39.46	At2g39020F	TCACCTCTCGTCCTTTCCAATCA
39.46	At2g39020R	GGTTGCTGGAAATGGAGAACG
39.46	At2g39010F	ATCAGTGGATCTTCTGGGTCGG
39.46	At2g39010R	CAACACAAACTGATGGTAAAATGCT
39.46	At2g39000F	GGCTGATTACCTTCCAAGAAAAGGA
39.46	At2g39000R	GCCACATTTGATACATAAGCGATC
41.08	At5g14850 R.T.Fw	AGAGGTAAGGAGGTTTTTCCACG
41.08	At5g14850 R.T.Rv	ACATAGACGACCACTGAGGATTGA
41.66	At4g27250 r.t. Fw	TCCCATCAAGCGTTCAAGTTCT
41.66	At4g27250 r.t Rv	CCCATTCTTTTGTTCACAGCTGA
41.66	At4g27260 r.t. Fw	GTGGTCTTGAGCATAGATTCCGA
41.66	At4g27260 r.t Rv	GACAAGGTGTGTCACTGCGTTCT
42.63	At1g25220F	ATCCTGTCTTCTTCAACCCAAGTC
42.63	At1g25220R	GGAAACTCTTGTGGGATTCGTAAG
42.63	At1g25230F	TGAGTGTCGTTTTTTATGATGTCTCC
42.63	At1g25230R	CGAGTCTAACATTTTCGATGTGTCC
42.63	At1g25240F	GATCAGCATAGAAAGTTTAACATGGAT
42.63	At1g25240R	GAAGTAAACTGGCTCGTCGATAACTA
43.62	At4g39640	GGTGGGAAAGAAATTGCTTATGATT
43.62	At4g39640	CCTCCATACATATTCTCAGTGGCTCT
43.62	At4g39650	TGAAAGATTTGCAGAATTACAACGT
43.62	At4g39650	TAATCGGTATCCAAGAATCTCCGTA
43.62	At4g39660	ACAAAGGCAACTCCTGAAGAGAGC
43.62	At4g39660	CAGTACGGAGGAGAGAGATCGC
47.15	At5g15500F	TGATGATACGTTTGTTGAGTTTAGGA
47.15	At5g15500R	GACCATAGGAACGAACAGAGGG
47.15	At5g15510F	CCAAGGCAGCTGAGATAGATTGTA
47.15	At5g15510R	TGCAACACAAACACACTTCATGTT
47.25	At5g66570 r.t Fw	ATTGACCTATGATGAGATTCAGAGCA
47.25	At5g66570 r.t. Rv	TCAATAGTAGGGCACTGGTTAGCC
47.25	At5g66580 r.t Fw	ACTTACCAAAAGAAAAACATTGCCG
47.25	At5g66580 r.t. Rv	TCTTCCCTTTACCGCAACCTC
47.64	At4g37290F	GGTCTAACAAAGACCGAGGAGAAGT
47.64	At4g37290R	TGAGTCTTTGATCGAACCAAGTGA
47.64	At4g37295F	GGAGGAAATCGATCAATCTAAAATTC
47.64	At4g37295R	CCATGTATCACGACCCCTAATAAGT
49.07	At2g16340 R.T Fw	GAGGGCAGTGGTTCCAATGAA
49.07	At2g16340 R.T Rv	AGTTTCGCTATCGGCGGA
49.07	At2g16365 R.T Fw	TGGGCGGCGTTATGATGAATATA
49.07	At2g16365 R.T Rv	AACTTGAACCCACTTTCTCTTTCCT
49.07	At2g16370 R.T Fw	AAGATAAACCCGGAGAAGAAACAAA
49.07	At2g16370 R.T Rv	GGATCATAGCCTGTGAGGTCAAAG
49.34	AT2G45650 R.T. Fw	TACTACGTGCAAGGTGAAGGGTCTT
49.34	AT2G45650 R.T. Rv	CCTTGGACGAAATTAGTCTCACCTG
49.34	AT2G45660 R.T. Fw	GAATAGAGAATGCAACAAGCAGACAA
49.34	AT2G45660 R.T. Rv	CTTCAACAAACCATTCCTTCTTTTG
49.34	AT2G45670 R.T. Fw	GAACTCGGAGAAGCTCTCAAAAAC
49.34	AT2G45670 R.T. Rv	GTACATTCCTCGAATCTCGTCCTT
49.41	AT4G38980 R.T. Fw	GGACAATGATGAGGAAGATGGGTT
49.41	AT4G38980 R.T. Rv	GGACAGTGGAAGACTCATCGAGG
49.41	AT4G39000 R.T. Fw	GTCTATACCGTCTGTTACCGTTCATC
49.41	AT4G39000 R.T. Rv	GAGAAAATATTCCATCCTGCTATGC
49.41	AT4G39010 R.T. Fw	GAGAGGACGATATGTACGATGATGAT

Table S6 continued: List of oligonucleotides used for mutant screening with real time PCR

49.41	AT4G39010 R.T. Rv	TGATGTAAGTAGTGGGCTCGGACT
49.44	At4g39000F	GTGCATAGCAGGATGGAATATTTTC
49.44	At4g39000R	CAATCACCGCACCTATAAGAATGTT
49.44	At4g39010F	GAGGACGATATGTACGATGATGATC
49.44	At4g39010R	GGAGCATTGATGTAAGTAGTGGGC
49.44	At4g38990F	ATGCTCAGAAGGATGGAACAGTTTC
49.44	At4g38990R	CGATCACTGCACCTACTAAAACGTT
49.53	At1g52180 r.t. Fw	AGGTGTTGCTTCTCTTACTGTTTCAGT
49.53	At1g52180 r.t. Rv	GAATCGCGTTGGAAGTTGGAA
49.53	At1g52185 r.t Fw	AGTTATCACGATTTAGCACATATCAATCTA
49.53	At1g52185 r.t Rv	GTTTCTTTCGGATTCTGGATTACAA
49.53	At1g52190 r.t Fw	AGGAATGGTAGGGTTAATGGTGTG
49.53	At1g52190 r.t Rv	TTCTCGAATCCTTTGCCAACA
49.53	At1g52200 r.t Fw	GCAGAAGTGATGGATGAAGGAGAG
49.53	At1g52200 r.t Rv	AGCCGGCATCATCAATAAGTACAT
49.53	At5g13330 r.t Fw	GTCAACCTTTTTCAACGCCTTCTT
49.53	At5g13330 r.t Rv	TCTTCTCTTCTTCTTCACGCTGCT
49.53	At5g13340 r.t Fw	AAGAGCAGATCTCCTGCACCAA
49.53	At5g13340 r.t Rv	CAATCGCAATTCTATGCTCGGA
50.02	At2g38280 r.t. Fw	CCATGGGAAAAAGAAGTCATATCTGA
50.02	At2g38280 r.t. Rv	GGATAGTGTGCAAATGGCTCTGTAT
50.02	At2g38290 r.t. Fw	AAAAGGTGGATG ATACATTAGCGG
50.02	At2g38290 r.t. Rv	CAACCCTGTCATTATTCCACCAAG
50.02	At2g38300 r.t. Fw	CGAGACTTCCACAGATCGAAATATT
50.02	At2g38300 r.t. Rv	TGGTTATGATTATAGCCTCGGAACA
50.36	At1g04360 r.t. Fw	CTCTCCTCGGGAGAATAGCCCT
50.36	At1g04360 r.t. Rv	ACGAAGTCATCATCGCCTCCTA
50.36	At1g04370 r.t. Fw	CTCATCCACTGCGGCTAATTCT
50.36	At1g04370 r.t. Rv	GAAGTTCATCCAAAACGCTATCGT
50.36	At1g04380 r.t Fw	CGCTTGACCTCACCAAAACATTTAT
50.36	At1g04380 r.t Rv	CCAGTTACCAGCAGACGAACTGTAA
50.46	At2g43610 RL F477	CAGAGGAAGGTACTGCTCACCAAG
50.46	At2g43610 RL R547	GACCGCGACCATAGTAGTTCTTTC
50.46	At2g43620 RL F682	AACCTAAATGTTCGTCCAGTCTTGG
50.46	At2g43620 RL R752	CATTCACCACCATTGATCTTCCTAG
50.54	AT2G34830 RL_FW	TCAGGAAAAGATGGAGTTCGATTAT
50.54	AT2G34830 RL_RV	TCTTGCAACAACTCCGGTCTATAG
50.54	AT2G34825 RL_FW	GAAAACAATTATGCAATCATTCGC
50.54	AT2G34825 RL_RV	GCTTCTGTTGTGCTCATGACGATAG
50.54	AT2G34820 RL_FW	GTATTTGCTTGAATCGCAAGCAA
50.54	AT2G34820 RL_RV	CACGGTACCAAACACACTTCCTCT
51.14	AT3G09250 RL_FW	TGCGATGCAATCACTATGGTC
51.14	AT3G09250 RL_RV	CCAGTTATGCCTTTGGCCC
51.14	AT3G09240 RL_FW	ATGTAGGAGTGAGCATATGGCCAA
51.14	AT3G09240 RL_RV	ACAAGCAACCTCTCGTGACCTTCT
51.14	AT3G09260 RL_FW	CTTCTTACGTCAATGCTAAATGCC
51.14	AT3G09260 RL_RV	TGTGAGTGACAAGGTAAGCCTCG
51.14	At3g09250 RL_Fw	GTTTTATCAGCCAATTCTCGGTTC
51.14	At3g09250 RL_Rv	CATAGTGATTGCATCGCAGCTAA
51.14	At3g09260 RL_Fw	ACATTAAGGACAAATACGCAAATCC
51.14	At3g09260 RL_Rv	GGCCCCTAGTTCTTCTCCATATC
51.70	At1g01460r.t.Fw	AAACAGCTCAAGCACTAACTTCGGT
51.70	At1g01460r.t.Rv	CTGATTCTCCTCTCCCAACCC
51.70	At1g01470r.t.Fw	ATTCGATTCCGATCTGTGAGATC
51.70	At1g01470r.t.Rv	GTATCTTCCCCTTTCCAATCTCC
52.18	At3g54830 r.t. Fw	TGACCAGAGCCTCTACATCGAGA
52.18	At3g54830 r.t. Rv	GTTCCATCGTCGTCTTCATCGT
52.18	At3g54840 r.t. Fw	TGTGCTTCTTCTCTTCCAGATAGGA

Table S6 continued: List of oligonucleotides used for mutant screening with real time PCR

52.18	At3g54840 r.t. Rv	CAGCTGGAACAGCATTCTCTGAAT
52.38	At1g14410 r.t Fw .	TCGCAACTCTTATCGACTCCTCTAA
52.38	At1g14410 r.t. Rv	AGTACTGAAGACGAAGAGAGAAAGCG
52.38	At1g14420 r.t Fw .	GTGCAGATGGTATGATCGATGCTAT
52.38	At1g14420 r.t. Rv	GTCGGTGAAATGACTGTTGGAAATA
52.38	At1g14430 r.t Fw .	TGTTTTCGCTTATAAAGCTGACGTG
52.38	At1g14430 r.t. Rv	AACATCATATAATACCCAGGTGGTGC
53.11	At5g43180r.t.Fw	ACTTCACTTGGCATTGATTCCTCA
53.11	At5g43180r.t.Rv	GTCTTATCGTCACCCTCTTTGATGT
53.11	At5g43185r.t.Fw	TCTATTGTAGGAATTTGTCGACGTTG
53.11	At5g43185r.t.Rv	ATAGGATGAATGCTGGGAGAATTT
53.11	At5g43190r.t.Fw	CTACGAGCATGTGTACTGTCTGTGG
53.11	At5g43190r.t.Rv	AGAATCTCAGGCCAATTGTAACAAC
53.11	At3g63030 R.T. Fw	GAAACAGGGCATGTCTTGTGAAGA
53.11	At3g63030 R.T. Rv	CCAAGTCCGAGAAGAATCGTAGTC
55.16	At1g31880 r.t. Fw	GATTGAAGAGGACGAGCCTGGT
55.16	At1g31880 r.t. Rv	CGACGTAGCTCTCTAGTCCCATCT
55.16	At1g31885 r.t. Fw	GACCTTAGCGGCTGCAGTTCTA
55.16	At1g31885 r.t. Rv	TTCCCACATAAACATCGCCTTT
55.16	At1g31910 r.t. Fw	ATCACATTCAATGCTGCGGAGT
55.16	At1g31910 r.t. Rv	TGCTGCAGAAGAACCTAAGCCA
55.53	At3g57660 r.t. Fw	GGAAGTTGGAGGACGCTATAGAGAC
55.53	At3g57660 r.t. Rv	ACCTGTGACATCATTGTGTATGCC
55.53	At3g57670 r.t. Fw	CCTCCTCTCCGAGAAGCCCT
55.53	At3g57670 r.t. Rv	GGTTGTGATGGTGGTCTTGTTGTT
55.53	At3g57680 r.t. Fw	TTGACAACAGTACCATCGGAAGAAG
55.53	At3g57680 r.t. Rv	CGACAAAGAAACCGCTACTGAAAC
57.31	At5g59340f	GTTCACTCAGATTTGACAGATTTGCC
57.31	At5g59340r	GGAATTGACGTTAATGTCCTCACC
57.31	At5g59330f	TCAAAAGGTGGTGTGGTACCG
57.31	At5g59330r	ATGGTTTGAGCCACACCGT
59.67	At1g05230 R.T.Fw	GCCACTTCTTTTTGGATCCCTGTT
59.67	At1g05230 R.T.Rv	TCTCGTCTCTGAGGAAGTCAAAGAC
59.67	At1g05240(50) R.T.Fw	TGAGACCAAAAACTACGTTCAGACTC
59.67	At1g05240(50) R.T.Rv	ATTGAAAGAAGAAAACACAGGAGGG
59.67	At1g05260 R.T.Fw	TCTGAGTTTGCCAAGTCAATGGAG
59.67	At1g05260 R.T.Rv	CAGCTGAACCAGTCTTGACATTGAT
61.36	At3g11130 R.T.Fw	AGAAGGATGTCATGTCACAACAGAA
61.36	At3g11130 R.T.Rv	GCGGGTAGAGCCAGTGGTAATAAT
61.36	At3g11150 R.T.Fw	GAGTACGCATGGAGAGTGTACCAAG
61.36	At3g11150 R.T.Rv	TGTATCTGTCAAGTGGATCCCTGA
61.36	At3g11160 R.T.Fw1	ATCTATGGGAGTCTATGGGAGTCAG
61.36	At3g11160 R.T.Rv1	TCGTTAGATAATCCCACATCTTTTC
61.36	At3g11160 R.T.Fw2	GAGGTCTGATCTATGGGAGTCTATG
61.36	At3g11160 R.T.Rv2	TAGATAATCCCACATCTTTTCCTCC
61.64	At3g07850 R.T.Fw1	CCATCGCAAGTGAAGTTGTCAGAC
61.64	At3g07850 R.T.Rv1	TTGTTGCTGATGTACCCTTAATGC
61.64	At3g07850 R.T.Fw2	GGCATTAAGGGTACATCAGCAAC
61.64	At3g07850 R.T.Rv2	GCACTCCTTTGCTGCACATTAG
61.64	At3g07860 R.T.Fw	GCATGTATGGTCAAACTTCTGTCTTT
61.64	At3g07860 R.T.Rv	CAGCATTATCGTCAAGCAACTTCTC
61.64	At3g07870 R.T.Fw	GTTGCCTAATTGGTTTCACACTGTTG
61.64	At3g07870 R.T.Rv	GGGTGTATCAAACCAAGACAATGTTC
61.64	At3g07880 R.T.Fw	GAAGGGAGCAAGTACAACTTGAAATT
61.64	At3g07880 R.T.Rv	TGGTGTACCTAAGACCAGAGACAATG
61.64	At3g07860 RL_Fw	CTTGGAAGCATGTATGGTCAAACT
61.64	At3g07860 RL_Rv	AACAGCATTATCGTCAAGCAACTT
61.64	At3g07870 RL_Fw	TATAGCATTGGGACTTACTTGCCA

Table S6 continued: List of oligonucleotides used for mutant screening with real time PCR

61.64	At3g07870 RL_Rv	GTTCTTCCAAATCCACATTGGTC
61.64	At3g07880 RL_Fw	CATACAACCACGTGATGCCAGAA
61.64	At3g07880 RL_Rv	TCTTGCAGAATATGATCCTCGAGC
62.40	At3g60047 R.T.Fw	CTTGGACACATGGAGAAGAAGCTTAG
62.40	At3g60047 R.T.Rv	CAAGCGAAGGGCTTACCTTATTTA
62.40	At3g60640 R.T.Fw	CGATGATGTCAACCATTTACGATG
62.40	At3g60640 R.T.Rv	TGTGTTTTCCCCACTGTAGGTAACATA
62.40	At3g60650 R.T.Fw	ATCCAAAGAAGCAGCCTAAATCAG
62.40	At3g60650 R.T.Rv	TGTAGTGGGCTGCATATAATCAAGAT
62.40	At3g60660 R.T.Fw	GGGACCAAGTTTGAAACTAGACAACA
62.40	At3g60660 R.T.Rv	CAAGGTGGCGAAGAACAGTAAGTATT
63.16	At1g08670f	GGTTCGATTGTGGTCTAAAGGTTC
63.16	At1g08670r	AAACATGTCGTCTTCAAGCAATCT
63.16	At1g08680f	CCATTTGATCTCCCCTATGATTCT
63.16	At1g08680r	CCTTGTAACGAACTCATGTCCAAG
63.17	At2g19600f	GTGGTTCGTGCAGTAGCTATTCCAG
63.17	At2g19600r	GTTATTCCACTCAAGCACATGAACA
63.17	At2g19610f	GAGTCTTACCTGAAATCTGACGCG
63.17	At2g19610r	GAACACTTTTTCCATCCTTGTGTCT
63.17	At2g19620f	TAACAGGAGACATGACTTAACTGATGG
63.17	At2g19620r	GGTCGCCAACAAAGATGAGTGTA
63.67	At5g65690F	CGACGAGCTTCACAGTCTCC
63.67	At5g65690R	GGCGCTGCCATCTCTAAGA
63.67	At5g65683F	TCAAGATCAACCGGAGGCAGT
63.67	At5g65683R	TGGTAGACACGACGAACGGA
63.67	At5g65687F	GGCTTCTTCAACTGTTTTGATCC
63.67	At5g65687R	CCTGCATTTTCCCATAAAGTGG
64.02	At1g07175f	TTATGCAATTCGGCTGTGTTTATG
64.02	At1g07175r	TCATCTTTTCGTTTTCTGAAAATGA
64.02	At1g07180f	GGCTACTATTGGGAGATACAAAGCTC
64.02	At1g07180r	CCATTGATATTCCTTTCCCTTCCT
64.02	At1g07190f	CTGGGAGACTCACTCTAACTGGAAA
64.02	At1g07190r	GGTCAGCTTCCGGGAATATTATAGT
64.15	At5g11530f	TATGTCAGCCGGATATTGAGTACAA
64.15	At5g11530r	TGATCAGAGTCACAACTTTTGGAAG
64.15	At5g11520f	GGACAACCTGAGCTCTCTTTATGCAT
64.15	At5g11520r	CAAGATGAGAGAAAACAGAACCACC
64.22	At2g19893f	ATTTCTTGTCTTCCTCTGCCTTTC
64.22	At2g19893r	CAATCGTTCATCTGGCAACTAATATC
64.22	At2g19890f	GCCTACTTGATCGAGGAATCCTTATC
64.22	At2g19890r	GCTACAAGCACAAGTCTCATCATGTT
64.47	At1g20490F	TTTCAAGGATATGCGTTGACTGAG
64.47	At1g20490R	TTTCAGACTCTCCTCCACAGATTCC
64.47	At1g20500F	AACCAAGAAGCTACAAATGAAACCA
64.47	At1g20500R	TCGATGTAGCAAAGATCTCCAGTT
64.47	At1g20510F	AAGGAAAACTGGGAGCAGCTTATC
64.47	At1g20510R	TGTATGGTGCTACTTGTTTCGCTAC
64.48	At1g70970f	GCAACCCTACCACCCGAAC
64.48	At1g70970r	CCGGGAATACATCTTTCCTCTTTAC
64.48	At1g70980f	AAGTGTAGCAAAGGCGAGTTTAGC
64.48	At1g70980r	CATTCTTAGGTAGTCCAGGCTTAAGC
64.59	At2g38300f	GATCAAACGTTGTCCTTGTCGCTAT
64.59	At2g38300r	TCATCTTTCATCAGTCGGCTCTTT
64.59	At2g38290f	CTGGAATGTGGTGTCGACTACTATCA
64.59	At2g38290r	CCTCAGCCATTCTCAATGGTATGA
64.77	At3g60966F	CTTTGATCGTCACAAGTCCATGG
64.77	At3g60966R	GTTTTCCGCCGCTCTCCAC
64.77	At3g60970F	CTCCAATGTCATATTTCGATTCAAC

Table S6 continued: List of oligonucleotides used for mutant screening with real time PCR

64.77	At3g60970R	GACACTTTGATCTGTGGATGCC
65.32	At1g63820 RL_Fw	GATCTCTACTGCAACAATCCTTATCTCA
65.32	At1g63820 RL_Rv	CTTTCAAAGAAACTAAGAGATTGATGATCAT
65.32	At1g63810 RL_Fw	ATTTATGTCGCCTGGAGATTTGAA
65.32	At1g63810 RL_Rv	GCTCGAAGTCCACCATTAGCTTATT
65.32	At1g63820	TGCATGAGGAAGAAGAAGCTTGTG
65.32	At1g63820	CATTTGCAACTGAGGCTGAGACT
65.32	At1g63810	ATTTATGTCGCCTGGAGATTTGAAG
65.32	At1g63810	GCTCGAAGTCCACCATTAGCTTATT
67.31	At2g45650f	AAAACCTTTCAAGACTTATGGGCAA
67.31	At2g45650r	GGAAATTCAGAATTGTTTGGATCC
67.31	At2g45660f	AAATTGAGCAGCTCAAGCAAAAGG
67.31	At2g45660r	TCCCCACTTTTCAGAGAGCTTCTC
67.31	At2g45670f	CTCGGAGAAGCTCTCAAAAACACAA
67.31	At2g45670r	AGCAAATGGTACATTCCTCGAATCT
67.51	At1g47813f	TCAATGGTGATACTCATAATGTTCAAGAT
67.51	At1g47813r	GGTTTCACAGGAGAAGCATCATC
67.51	At1g47810f	CCACTATATGAGCCAAACTCACTGG
67.51	At1g47810r	TCCCCAACTAGAGCTTCAAATTT
67.74	At3g04570F	CCGGTGTATAATATGCCGGGA
67.74	At3g04570R	TGGCCGCTCATCTGTCCTC
67.74	At3g04560F	AGAGAACAAATGAAGATCCAGATGC
67.74	At3g04560R	AATCAAAGCTTCTCTTTTTGACCTGA
67.74	At3g04580F	CTTAGGCATGTGTAGAAAACTTGCA
67.74	At3g04580R	TGTCCATGCGATTTAGGTGATATC
68.24	At1g71000 RL_Fw	CAGAACCAATCAATGCAAATGAATT
68.24	At1g71000 RL_Rv	GCGGTAGAGACATCTGTGAATGC
68.24	At1g71002 RL_Fw C-Ter	TGATCATAAGAAGGATGGATGAATG
68.24	At1g71002 RL_Rv C-Ter	TCTCGATCTCTCCTCAAAACCCTA
68.24	At1g71002 RL_Fw N-Ter	TCAGAAATCAAAAGTTTCGATCGTC
68.24	At1g71002 RL_Rv N-Ter	AACGTACATGGTTTATCGGTTTGTT
68.41	At5g41740 RL_Fw	TTTGAGTTCAGCCACACTAAAATCG
68.41	At5g41740 RL_Rv	GCTCCCTCTGTCATGATCTGTACAC
68.41	At5g41730 RL_Fw	CAAGATCATCGCCAAAAGTATCAC
68.41	At5g41730 RL_Rv	TCGATTATCCTTCCTTGCTTTAATG
68.41	At5g41750 RL_Fw	GGTGGAGCTTCCCTTTTCTATTTT
68.41	At5g41750 RL_Rv	AGCATAGAGCAGTTTTCAACATTCA
70.24	At5g59330F	TGGTACCGGATCTGTGTTGTG
70.24	At5g59330R	TGTAGGCATCGGCATGGTT
70.24	At5g59340F	TTCACTCAGATTTGACAGATTTGCC
70.24	At5g59340R	TGGAATTGACGTTAATGTCCTCAC
70.70	At5g49120F	GAAGAAAGCTTGAGAAGAGAGTATTGTT
70.70	At5g49120R	TTTCTCGGATCTCGACGGTAA
70.70	At5g49130F	TCAGACCATTAGCTGTGGAATCTTAC
70.70	At5g49130R	CGTAGAAATTAATCTTGGCCCCTAT
70.70	At5g49140F	TGAAGAAGTGGATGTGATACGGTT
70.70	At5g49140R	ATTTGTACACCACATTCCGTAATCTC
71.12	At2g30200F	AGGATCCCAGTGATCTCGAATGTC
71.12	At2g30200R	CACCTGGCGTGCAAGTATCTT
71.12	At2g30210F	CTTATTTGACCCACCAGAGAGGAAC
71.12	At2g30210R	CGACGAACCGAATTGCCAC
71.12	At2g30220F	CTTGTCCGAATCATTCGGATCACT
71.12	At2g30220R	TAGGTATTTGTAAGCTGCCTCGGAG
71.34	At3g03470F	TTATTTGTTGGCTCTGATGTGTTTC
71.34	At3g03470R	CATTTGGTACTGAGCTTCTTCGATC
71.34	At3g03480F	TTCAGTGACCGCCCTAATGG
71.34	At3g03480R	CCGATATTATGTAGTTCCCTGATGC
71.34	At3g03490F	AGGAGCTTGGCGCAAGATATCC

Table S6 continued: List of oligonucleotides used for mutant screening with real time PCR

71.34	At3g03490F	GTGAATATCGTTTGTAGTCTTCTTTGCTT
71.36	At3g04100f	TCGAGTTACGTGATAACTTATGCGA
71.36	At3g04100r	AGATGAACCACCACAGTCTTTTCC
71.36	At3g04110f	AGCTGCTCATGTTCTCGTACTAGC
71.36	At3g04110r	TGGTGCATAGTCGTGATACCTCTT
71.72	At5g67411f	CTACATGTGGATTCCTTACGATACCG
71.72	At5g67411r	CATACCACTGTCTTTGCTCGCTAC
71.72	At5g67420f	ACCGTCGATGAACTCAGAGGAC
71.72	At5g67420r	ACACATCACCCTTGTCACAAAAC
72.14	At1g02930f	TTTGAAGATGGAGACTTCAAGATTTT
72.14	At1g02930r	TCTGAGAATTCATGAGCTATGTATTGAGTAA
72.14	At1g02940f	TTGGCTAGTAACAGCTTCACAATG
72.14	At1g02940r	TGTGTGTCCATAAGGTATTGTATATTAGGA
72.14	At1g02950f	CGAAGAAATTGTTCGAAAAAAGATC
72.14	At1g02950r	CCAAGCCTCTCTACTCGTAATCTCAT
72.29	At3g59580F	ACAAGAATCGAGCGAGGTAATGGA
72.29	At3g59580R	GATGAATCCGACATGCTACTTGATAT
72.29	At3g59590F	GTGTCCACCTCACTGCTCTCG
72.29	At3g59590R	GGTGAATGGTTAGGAAACGAATAATA
72.29	At3g59600F	GCAGAGTTATATGTTTCGTTCGGG
72.29	At3g59600R	CGAAGTGAGAAATATGAGCTGGATC
72.67	At4g37580F	TTACATGTGGCCAAAAACTCGAT
72.67	At4g37580R	TTTAGTGTAAAGAGGCTTAACGACATC
72.67	At4g37570F	GGAAAATCTCGGGTTTACG
72.67	At4g37570R	CATAAACCCGTGATTTTCC
72.67	At4g37560F	AAAGTTCCAACAGGAGCTCGGATAG
72.67	At4g37560R	GGAAGTTTTCCATCATATGTGCTCTT
73.63	At1g05660 RL_Fw	GAAAACCCAATCATAATCGACCAA
73.63	At1g05660 RL_Rv	ACCCCAGAATACTCATTAGGGCA
73.63	At1g05670 RL_Fw	TGTCTGTTGATTCTTGCAACGTG
73.63	At1g05670 RL_Rv	TGATTGCTGTTGCAGTTTTATAGCA
73.63	At1g05650 RL_Fw	CGGATTCGTTAGAAACGTCTTCTTT
73.63	At1g05650 RL_Rv	GGTCGATTATGATAGGGTTTTGGAC
75.66	At3g13900 RL_Fw	GCTTACAGTTCCCAGCACTGTACC
75.66	At3g13900 RL_Rv	CCCGATGATTCTGTACCAGTCAA
75.66	At3g13910 RL_Fw	AGACGCAGTGAAAGGTTACAACAG
75.66	At3g13910 RL_Rv	AATCGTTCGTTTGACTCTGATCG
75.66	At3g13890 RL_Fw	AAACTCAACGTGTTTGTGACACCA
75.66	At3g13890 RL_Rv	GCCTTCGAGAACTTCTCCACATAA
79.23	At3g27510 RL_FW	AATAGCTCTTCACGACCACTTTGC
79.23	At3g27510 RL_RV	TCTGCCCTAAGAATGAAGGAACC
79.23	At3g27520 RL_FW	GAGAAGATGCTTCTAAGTCGCCGT
79.23	At3g27520 RL_RV	CGCTTCTTGTCTTTCCACCGTAA
79.23	At3g27503 RL_FW	GTAATGGAACATGTGGGCCTAACA
79.23	At3g27503 RL_RV	GGGAGAAGGAGGATTTTTCCAAA
79.23	At3g27500 RL_FW	TTTGTGAAAGAGAGACCAATCCAA
79.23	At3g27500 RL_RV	AACATGTAAAGTGACCCCGCA
81.21	At5g46540 RL_FW	AGGGACAATATTACCAATCGTCCA
81.21	At5g46540 RL_RV	GTCGCATAGGATACCGGAAACTA
81.21	At5g46550 RL_FW	AGGAACAGATGGAAAGAGCACAA
81.21	At5g46550 RL_RV	CAACCTTAGCAGCTCTCATCTCG
81.21	At5g46530 RL_FW	CACTTTTAGCCATGGGATTTGGT
81.21	At5g46530 RL_RV	CATTGTCCTCAGACGAGCATTTC
81.77	At2g37440 RL_Fw	TTATCGCCAAATGAGCTCGAG
81.77	At2g37440 RL_Rv	TCTGAGAGGCCTGAGAAGCTG
81.77	At2g37450 RL_Fw	AGATGTACCTCGGAAGGGCTCTT
81.77	At2g37450 RL_Rv	TTTGCCCCATATCACAAGGTATAG
81.77	At2g38290 RL_Fw	TGGAATGTGGTGTCGACTACTATCA

Table S6 continued: List of oligonucleotides used for mutant screening with real time PCR

81.77	At2g38290 RL_Rv	TCCTCAGCCATTCTCAATGGTAT
81.77	At2g38300 RL_Fw	GATCAAACGTTGTCCTTGTCGCTA
81.77	At2g38300 RL_Rv	TCATCTTTCATCAGTCGGCTCTTTT
82.20	At3g01080 RL_Fw	TTGATAACTTCCCTGACGTCTTGA
82.20	At3g01080 RL_Rv	CGAGGTGAGAAGATGAGTAAACGAC
82.20	At3g01070 RL_FW	ACCTAACAAGAACAGTGCAAGAAGAA
82.20	At3g01070 RL_RV	CTACTGGAATAAGAAACTGATAGGCAAA
84.27	At5g65630 RL_Fw	GAGATCGAAAGAGATGGTACTGCA
84.27	At5g65630 RL_Rv	CTGGAACTGCCTGAAGAACTAGAAC
84.27	At5g65640 RL_Fw	AATAGCAACAATTCACATCACTCTAAGCT
84.27	At5g65640 RL_Rv	ACCAGAGGTTCGTTCGCATTAA
84.63	At1g61270 RL_FW	AGAAACAACACCTCAGTGCCATT
84.63	At1g61270 RL_RV	TGACCTGCATACGCAAAAGC
84.63	At1g61260 RL_FW	AGATCTGGTTGAATCCGATGTCAT
84.63	At1g61260 RL_RV	TCTCAATCGGAGGAAGATTCTCC
84.63	At1g61280 RL_FW	ATCAAGAAGTACACGGACCAAAAAC
84.63	At1g61280 RL_RV	GGCAACAACAATCGAAATAGATCC
84.64	At5g59730 RL_FW	CCGAAACTTAGGGACCAGATTAAA
84.64	At5g59730 RL_RV	CGGTAAAACTCGGTACAAACCAG
84.64	At5g59740 RL_FW	GGCTGAGCCAGAATTAGTAAACGG
84.64	At5g59740 RL_RV	CGGAAACTGCGAATACACCTTTC
84.75	At4g30350 RL_Fw	CAGCGAAAAACGTTGCTTCTATC
84.75	At4g30350 RL_Rv	GTCTTAACTCCCATCCTTTGTTCAC
84.75	At4g30360 RL_Fw	TCCTAGAACAAAGATGAACCTTGGA
84.75	At4g30360 RL_Rv	ACACCGCGTCTTGTATTAGCTGC
86.25	At5g46980 RL_Fw	AAATACCTCAAAACACTGGATTACGAGA
86.25	At5g46980 RL_Rv	GCTCTTGGTTTACCTTTTGCCAT
86.25	At5g46990 RL_Fw	AATTGCGGGATTTCGATAAAGTC
86.25	At5g46990 RL_Rv	CATTTCACACTCTCTCGGAACAACT
86.56	At3g02000 RL_FW	GATCATACAAGATGAGCAGCTTAGGA
86.56	At3g02000 RL_RV	GGACTCTATACGAAGCAGACCAGTG
86.56	At3g02010 RL_FW	TAGGCGTCTTAACTGCATGTAGTCA
86.56	At3g02010 RL_RV	AGACATTGCCTGGAAATATTCTGTT
86.56	At3g01990 RL_FW	CGCGAGAGGGTAAGGCTAAAGAC
86.56	At3g01990 RL_RV	GATTTTTGATTCAACCGGGTTTC
86.71	At5g11320 RL_Fw	GACTGGGTATAAGAGTAATGTACCCGATT
86.71	At5g11320 RL_Rv	TTGGCATTCCTTCTTTTGTGAAAA
86.71	At5g11330 RL_Fw	ACTCGGAAACAGTTAGTGGGATTAA
86.71	At5g11330 RL_Rv	CACCGAGATCAAAATACAAGCATT
86.76	At5g28370 RL_Fw	TTGTTGCAGCGTCAGTAATGTAGAG
86.76	At5g28370 RL_Rv	GAGAACAACCAGCTTCCAACATTT
86.76	At5g28380 RL_Fw	GATTCCGAGACGGTTGGCATAATA
86.76	At5g28380 RL_Rv	CGCTCAAACGGGATAGTCTTTTC
87.09	At1g62886 RL_FW	TTTGTGAAGCCATATGCTGAGGAG
87.09	At1g62886 RL_RV	GCGTAAGAGTTTTCGTCCCTGAA
87.09	At1g62880 RL_FW	GGAGAGGTCTGGACATGGATTATTT
87.09	At1g62880 RL_RV	AGCTGATAGACGATGAGTCCGAG
87.17	At1g30840 RL_FW	ACGTTAAGCTCTGTTTTGTTGGCT
87.17	At1g30840 RL_RV	ATACTTGGTTTTGGTTAAACCGGAC
87.17	At1g30845 RL_FW	AGAATCGTTATCGTTTCGTTGGTT
87.17	At1g30845 RL_RV	CGACTTACTGAGGATAAGAACTCCG
87.17	At1g28300 RL_FW	GTCCAAGTGAGTTGGTTGATTCAA
87.17	At1g28300 RL_RV	CGGAGATCTGCACGTTATTTTTC
87.17	At1g28320 RL_FW	GGGAAGTAAATTTTGGGAACAGAAG
87.17	At1g28320 RL_RV	CAGACGAATAATGATTTCGTGGAG
87.17	At1g28307 RL_FW	GCCGAGCTGGGCCAG
87.17	At1g28307 RL_RV	AACATCTTTTATGGGTCGAGGC
87.42	At1g15757 RL_Fw	CGTGTTTGCCTGAAGTCGGA

Table S6 continued: List of oligonucleotides used for mutant screening with real time PCR

87.42	At1g15757 RL_Rv	GGCCTGTTGTAAACTTCATTCTTGA
87.42	At1g15750 RL_Fw	AAGATCAAGCTCACTTCCTCGTTGT
87.42	At1g15750 RL_Rv	GCATTCG AGCTTGGTAGTTTCGTA
87.42	At1g15760 RL_Fw	GGAACAAGAGATCTGTGATGCCG
87.42	At1g15760 RL_Rv	TGGTCAACAACAACCTCTCTTGCT
87.58	At5g10180 RL_Fw	ACGGTTATAATCTGGCTATCAAACG
87.58	At5g10180 RL_Rv	TGAAGTGGATTTGACTCGTCCTC
87.58	At5g10190 RL_Fw	TCAAACGGATATCTGTAATCGACG
87.58	At5g10190 RL_Rv	CACGAATTTTTCGGTATATGCCA
87.59	At2g41860 RL_FW	ACATTCAAATATTGATGGACGCG
87.59	At2g41860 RL_RV	TGGCTACAAATTCATTGACGTCTAA
87.59	At2g41870 RL_FW	ACAACGAACTAATGACTCGTGACGT
87.59	At2g41870 RL_RV	CGTGGATTCTCATCAACTCGTCT
87.59	At2g41850 RL_FW	ATCACGGGATTGAGAAGTTTTGTAA
87.59	At2g41850 RL_RV	CCAAAATCAGAAACACTAACGGTG
88.05	At5g35770 RL_FW	TCACCCTCTCAGACCTCTACTTAGC
88.05	At5g35770 RL_RV	GTTGTTTAAAAGAAAAAGCCGGAC
88.05	At5g35760 RL_FW	TGCAAAACTCATGAAGCCTCTATC
88.05	At5g35760 RL_RV	CCCTCCACATCTCCTTGATTTTAG
88.36	At5g22500 RL_FW	CAGGCAATATATCATGTTGGTTCGT
88.36	At5g22500 RL_RV	GCCGTGAAATCGTGAAGTTGTC
88.36	At5g22510 RL_FW	GGCCAAGCTCCTCTTAGCAAATC
88.36	At5g22510 RL_RV	TTCTCAAATCCGAATCCTCTTCAC
88.36	At5g22490 RL_FW	TTTCTCACCGGATTGTGTCTTTG
88.36	At5g22490 RL_RV	CGTCGTTGATAGTCATTTCCATAGC
89.01	At3g60160 RL_FW	TGCTTGTGCCATTCATAATATTGG
89.01	At3g60160 RL_RV	CATGCCATCCAGTAATTGCTAGC
89.01	At3g60164 RL_FW	CCTCCGTTGCTGTCTTCTCTGTA
89.01	At3g60164 RL_RV	AGTGCGGTGGCGATACCAT
89.29	At5g55880 RL_FW	GTTGAATACATGGATGCTGACGA
89.29	At5g55880 RL_RV	GTAAGATCACCAACGATGGAGGAT
89.29	At5g55890 RL_FW	TACCACGCATGTTTTTCCCTAACA
89.29	At5g55890 RL_RV	GGCTAGATTGAAATATCCGTCTTCG
89.68	At3g12900 RL_FW	GAAGTACAAAAGTGCGGAGCATAG
89.68	At3g12900 RL_RV	TGAAAATGGGCACAGAGACTCTT
89.68	At3g12915 RL_FW	AGGTCAGCTAAGGGCTGCAA
89.68	At3g12915 RL_RV	ACATCATGTCCCAATGGTCAAAC
89.68	At3g12910 RL_FW	TGGCCCTACACACTCTAACCCTAC
89.68	At3g12910 RL_RV	AATATCCGGCATAACGTCCATACT
89.68	At1g79110 RL_FW	GGTTACTACGCCGATGGTGCTAC
89.68	At1g79110 RL_RV	CTTGAGCGTTTTCTCGTTGGTAAAG
89.68	At1g79120 RL_FW	ACGAGAGGATAGGTTATCCCGTG
89.68	At1g79120 RL_RV	TTGTGATCCTGACCCTCTTTCTG
90.36	At3g10114 RL_Fw	AAGAGAGCACAAGCGAGG
90.36	At3g10114 RL_RV	CGCCGTAAGCTAACCCTAATAAAGC
90.36	At3g10116 RL_FW	TTAGAGAAGAACATGACGACCTTGC
90.36	At3g10116 RL_RV	TTAGATTAGCTCGGCCAACATCA
90.36	At3g10113 RL_FW	TCAACACCCATTTCAAGCATTACTC
90.36	At3g10113 RL_RV	GGATGGTTTGTGAGATTCTTCCG
90.36	At1g01610 RL_FW	TGTTGCTCTGACCCGAGACC
90.36	At1g01610 RL_RV	CCAAATCACCTTTCTCGAGGAGTT
90.36	At1g01620 RL_FW	AGAAGAGGATGTTCGAGTGGGA
90.36	At1g01620 RL_RV	CCGTCTGAGCCGATGTACCTA
90.36	At1g01600 RL_FW	TCCGGGACATAAAGTGGAGC
90.36	At1g01600 RL_RV	ACAGATTCACCAGAAGTCCGTTC
90.46	At5g09530 RL_Fw	CACTAATGAAGAAGAGTCTCTCTGCTG
90.46	At5g09530 RL_Rv	CAATGCGATAAGACATATGATCAGAA
90.46	At5g09520 RL_Fw	AATTTCCCAAGTTCGATTTTCCA

Table S6 continued: List of oligonucleotides used for mutant screening with real time PCR

90.46	At5g09520 RL_Rv	AAACGCCGGCACTTTGGTT
90.46	At5g09540 RL_Fw	GATACCGACATTTCCGGAGCTAC
90.46	At5g09540 RL_Rv	GCAACTCTAGCTTCGGAAATTCG
90.54	At3g15500 RL_FW	GGTTCGTCGACATCTTCTTCGTC
90.54	At3g15500 RL_RV	TCTGTTGTCAATCTCATGCAACG
90.54	At3g15510 RL_FW	AGATCACGACAATAACGCTGACGT
90.54	At3g15510 RL_RV	TTCGCCATCATCATAGGACCTG
90.63	At1g77200 RL_Fw	AGTTACGATGAGTCATTGAGCGAG
90.63	At1g77200 RL_Rv	AATATACCAAGGCGAACTAGGTGG
90.63	At1g77210 RL_FW	CGTCGCTTATCTCGTCTACCATTAC
90.63	At1g77210 RL_RV	CAGCTGAATACATAGACATGATCGC
90.63	At1g77180 RL_FW	GAAGAGCAAGATCACCAGAGACAGA
90.63	At1g77180 RL_RV	CAGTAGATGCCATTCCAAGAGCA
90.64	At2g45050 RL_FW	CAAGAAGCAAACGATCAAGAGCT
90.64	At2g45050 RL_RV	GCTCGGATTCCAGTGGCA
90.64	At2g45040 RL_FW	TATCGAAGCTCGAAATCCATCTCA
90.64	At2g45040 RL_RV	CGTTTGATTTCCGGTATAGAGACGT
90.64	At2g45060 RL_FW	CTTCTCGATTCCGCACTCAAATA
90.64	At2g45060 RL_RV	CTTTGTCAAAGCGTTTCTCTCTTG
94.12	At3g63540 RL_Fw	GTGACTGTGATCAGGCTGTGCTT
94.12	At3g63540 RL_Rv	GACGAGACAGGGCAGGTCTACTAT
94.12	At3g63530 RL_Fw	CTCTTAACTACGATCATGCCGCAA
94.12	At3g63530 RL_Rv	TGCATTGGTATTCATGGTCCAGTA
94.31	At4g40010 RL_Fw	AGTGCAAGCATCTCTTGTCTCGTAT
94.31	At4g40010 RL_Rv	TTCGATTTCCGGTACAGTTATTCTC
94.31	At4g40020 RL_Fw	TGGTGAAGGAAACTCTCCAAGTTC
94.31	At4g40020 RL_Rv	TGGCGTCTCTGCTATATCAAAGATC
94.31	At4g40000 RL_Fw	AACTAGCCTTGGGATCTTGCGTTA
94.31	At4g40000 RL_Rv	CACTTTGAATTGTTCTGACCCGA
94.31	At4g40011for	AATAATATTGACGGCCCAACAGTC
94.31	At4g40011rev	CATACGTACAGATACAGTCAGCACCA
94.48	At1g75440 RL_FW	AGAACATTATCCTATGGAATCGCC
94.48	At1g75440 RL_RV	TGTAAATGTGAGGATGCAGAGGAG
94.48	At1g75450 RL_FW	ATGAACAAAGACAAATGGGACGAG
94.48	At1g75450 RL_RV	GCCACCAGATAGAAAACTTCCTCA
95.19	At3g01870for	TTTAATTGTTAACCATGTGTGTTGGA
95.19	At3g01870rev	TGCTCACTGCTGTAGAATTCACATC
95.19	At3g01860for	GTTCGTGTGTCTCCTCTTCTCAAA
95.19	At3g01860rev	GGTTTAGAGATGAAAGAGGCAAGC
95.19	At3g01880for	GCAGAAACTGGAACTAACCCTTTG
95.19	At3g01880rev	GGATATTTGAATGGATACCCTAAAGAATT
95.31	At4g11911for	TGGTCTCATTCTCATGGCTGC
95.31	At4g11911rev	TTGGTCAATAGATTCCCTCCG
95.31	At4g11910for	CTGATGAGTGCAAGTGTTGCTTTC
95.31	At4g11910rev	CGGTATGTTGATAATGACGATGAGA
95.31	At4g11920for	TCAGCTGAATCTTCCACCGTC
95.31	At4g11920rev	CGATTAAACGGTTGATTCGTGAC
95.36	At1g11920for	TACGGAAGAGCAGAGCGGTTT
95.36	At1g11920rev	GGACCGGCGGATGAAGTAAG
95.36	At1g11915for	AATTTCCAGTGACAATGAAGTGCCT
95.36	At1g11915rev	CGTGTCCTCGTATTAAACCCACAA
95.36	At1g11925for	GTTTGCTGCAACGGATATTGTGTAA
95.36	At1g11925rev	CGATCTCCCTTTCTTGCACTTCTT

Table S7: List of oligonucleotide primers used for quantitative expression analyses with real time PCR

gene	primer	sequence (5'-3')
CHS (At5g13930)	rtAt5g13930for	TGTTGAGCGAGTATGGAAACATGTC
	rtAt5g13930rev	TGACTTCCTCCTCATCTCGTCTAGT
FLS (At5g08640)	rtAt5g08640for	GATTTAGCTTTAGGTGTACCGGCTC
	rtAt5g08640rev	GGAACTTCGTTAGGAACAAGAAGAGT
DFR (At5g42800)	rtAt5g42800for	CTCCAAATTTCTCAGGCCAAAATAC
	rtAt5g42800rev	GATTCTCATCAACACCTTCAAACG
CHI (At3g55120)	rtAt3g55120for	AATTAGGGCTTTATACGGACTGTGA
	rtAt3g55120rev	ATGTTTCTTCCTTGAAGATCTCCAA
F3H (At3g51240)	rtAt3g51240for	GAGTCTCTTACCAATGCATGCGTC
	rtAt3g51240rev	TGAGGGCATTTTGGGTAATAATTAA
F3'H (At5g07990)	rtAt5g07990for	GAGAGCTGTGAGATCAACGGCTAC
	rtAt5g07990rev	GCTATGGCCCATATGTTCGTCAAT
PAL1 (At2g37040)	rtAt2g37040for	GTGAGCTTCATCCTTCTCGCTTCT
	rtAt2g37040rev	TGTGTAGACTTGTTCACGGTCTACAA
PAL2 (At3g53260)	rtAt3g53260for	GTCACAAGCCATGTTCAATCAGCT
	rtAt3g53260rev	CGAGACGAGATCAAACCAAGAGAGT
MYB13 (At1g06180)	rtmyb13for	AATTGGGAGGGCTCACTAGATAGAAA
	rtmyb13rev	AGATGGTCAAACCAAAACTCCATG
MYB20 (At1g66230)	rtmyb20for	ATATGATGAGCCTGTGGGAAATTAG
	rtmyb20rev	GGAAGAACTCTCATCATTAAGCAGC
MYB11 (At3g62610)	rt2_myb_11_for	ATTCGGTTTTGTCGAGGCTGTTAG
	rt2_myb_11_rev	GTCGTTGGAGTCCCTTTGTCTGAT
	RT_MYB11_At3g62610_FWD	TCGATTCTTGTTTTCTATTGGACGAC
	RT_MYB11_At3g62610_REV	TTTTACGTTGTTCGACTCACATGA
MYB12 (At2g47460)	rt2_myb_12_for	ACGCAGACTTGGGAGAACGAGTAG
	rt2_myb_12_rev	TCGTTTTACGAGTTTTTGTTCTGTGG
	RT_MYB12_At2g47460_FWD	TGGAGTCGTTTCTCAACTACGACC
	RT_MYB12_At2g47460_REV	ACAATCCCAATCGATAAACTCATC
MYB111 (At5g49330)	rt2_myb_111_for	GAGGGAAGTTTTTTAAGTTCGAATTC
	rt2_myb_111_rev	GAATTACATAGACCAACCCACCAAT
	RT_MYB111_At5g49330_FWD	GGTTGATGAGGAGTTGCTTGATTG
	RT_MYB111_At5g49330_REV	GATCCCATAGATCATCACTTTGACA
MYB28 (At5g61420)	RL_MYB28sh_Fw	TCCCTGACAAATACTCTTGCTGAAT
	RL_MYB28sh_Rv	CATTGTGGTTATCTCCTCCGAATT
MYB51 (At1g18570)	MYB51_RL_Fw	CTACAAGTGTTTCCGTTGACTCTGAA
	MYB51_RL_Rv	ACGAAATTATCGCAGTACATTAGAGGA
Actin (At3g18780)	Act_Foward	ATGGAAGCTGCTGGAATCCAC
	Act_Reverse	TTGCTCATACGGTCAGCGATG

Table S8: List of oligonucleotides (for semiquantitative PCR and T-DNA insertion line control PCR)

gene	primer	sequence (5'-3')
At3g09260	WiscDsLox461-464F2for	AGGAAATATTCTTGGCCATGC
	WiscDsLox461-464F2rev	TTGATTAATATGGCCTGGCTG
At3g62610	myb11_ln_for	CCCAAAAATGCCGGGCTAAAGAGATG
	myb11_ln_rev	TTCGTCAATATCCAACGGTTCTCCAAAGT
At5g49330	myb111_ln_for	ATATTCAGACCAATGGTGAAGGTT
	myb111_ln_rev	CCCATAGATCATCACTTTGACAAG
At3g18780	Actin lg F	TAACTCTCCCGCTATGTATGTCGC
	Actin lg R	CCACTGAGCACAATGTTACCGTAC

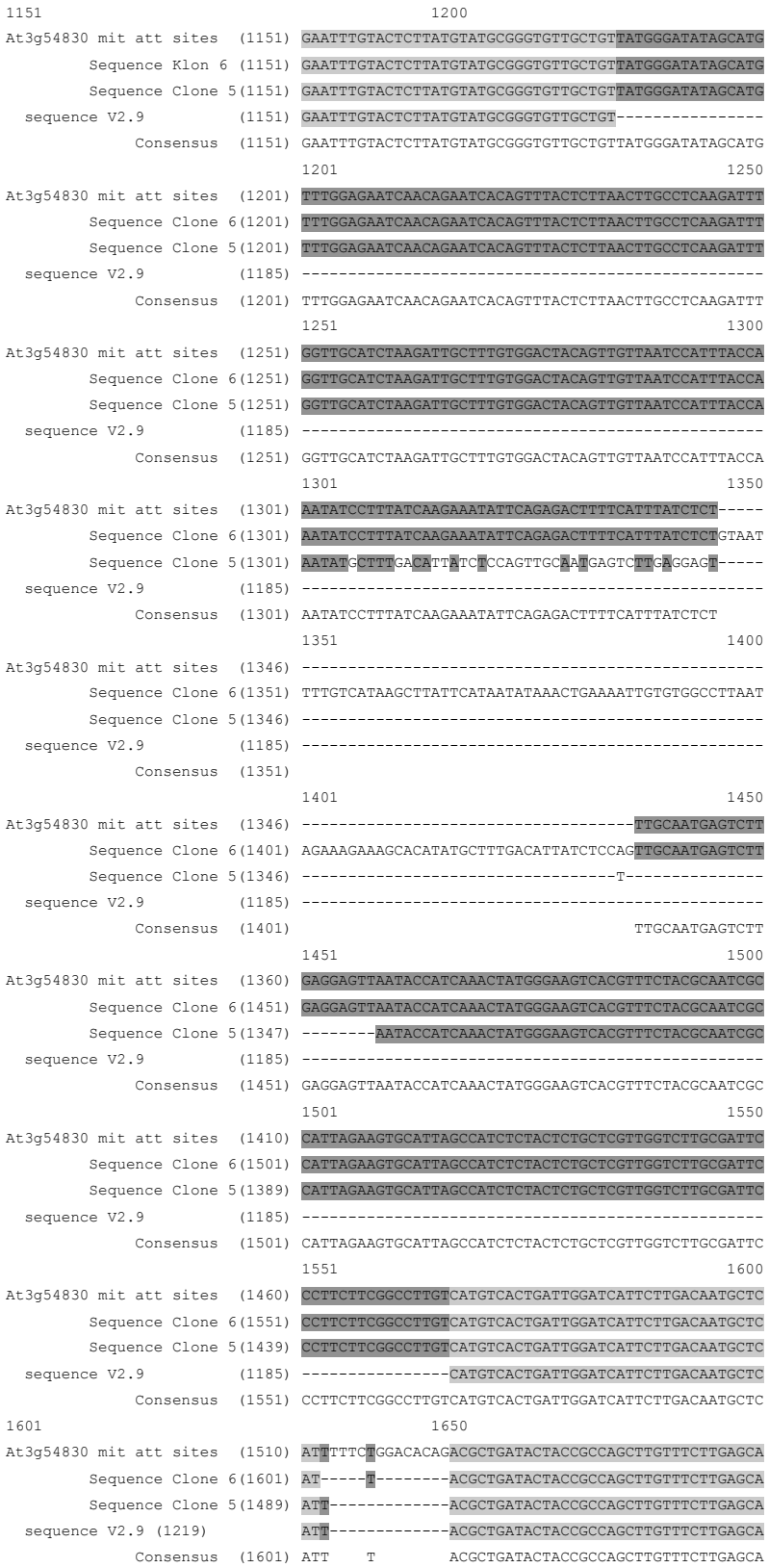


Fig. S7: 3' end of the At3g54830 sequence of clones 6, 5 and 2.9.

6 References

1. A. M. Smith *et al.*, *Plant Biology*. (Garland Science, 2010).
2. D. J. Kliebenstein, Secondary metabolites and plant/environment interactions: a view through Arabidopsis thaliana tinged glasses. *Plant Cell and Environment* **27**, 675 (Jun, 2004).
3. M. Wink, *Functions and Biotechnology of Plant Secondary Metabolites.*, Annual Plant Reviews (John Wiley & Sons ed. 2, 2010), vol. 39.
4. J. C. D'Auria, J. Gershenzon, The secondary metabolism of Arabidopsis thaliana: growing like a weed. *Current Opinion in Plant Biology* **8**, 308 (2005).
5. Y. J. Surh, J. K. Kundu, H. K. Na, Nrf2 as a Master Redox Switch in Turning on the Cellular Signaling Involved in the Induction of Cytoprotective Genes by Some Chemopreventive Phytochemicals. *Planta Medica* **74**, 1526 (Oct, 2008).
6. M. B. Sporn, Approaches to prevention of epithelial cancer during the preneoplastic period. *Cancer Res* **36**, 2699 (Jul, 1976).
7. L. W. Wattenberg, Chemoprevention of cancer. *Cancer Res* **45**, 1 (Jan, 1985).
8. W.-S. Jeong, M. Jun, A.-N. T. Kong, Nrf2: a potential molecular target for cancer chemoprevention by natural compounds. *Antioxidants & Redox Signaling* **8**, 99 (2006).
9. K. W. Lee, A. M. Bode, Z. Dong, Molecular targets of phytochemicals for cancer prevention. *Nature Reviews. Cancer* **11**, 211 (2011).
10. Y. J. Surh, Cancer chemoprevention with dietary phytochemicals. *Nat Rev Cancer* **3**, 768 (Oct, 2003).
11. K. Saito, F. Matsuda, Metabolomics for functional genomics, systems biology, and biotechnology. *Annu Rev Plant Biol* **61**, 463 (2010).
12. M. Jahangir, H. K. Kim, Y. H. Choi, R. Verpoorte, Health-Affecting Compounds in Brassicaceae. *Comprehensive Reviews in Food Science and Food Safety* **8**, 31 (2009).
13. T. Rauhut, E. Glawischnig, Evolution of camalexin and structurally related indolic compounds. *Phytochemistry* **70**, 1638 (Oct-Nov, 2009).
14. J. Glazebrook, F. M. Ausubel, Isolation of phytoalexin-deficient mutants of Arabidopsis thaliana and characterization of their interactions with bacterial pathogens. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 8955 (1994).
15. L. Browne, The camalexins: New phytoalexins produced in the leaves of camelina sativa (cruciferae). *Tetrahedron* **47**, 3909 (1991).
16. J. Tsuji, E. P. Jackson, D. A. Gage, R. Hammerschmidt, S. C. Somerville, Phytoalexin Accumulation in Arabidopsis thaliana during the Hypersensitive Reaction to Pseudomonas syringae pv syringae. *Plant Physiol* **98**, 1304 (Apr, 1992).
17. E. Glawischnig, Camalexin. *Phytochemistry* **68**, 401 (Feb, 2007).
18. I. Ahuja, R. Kissen, A. M. Bones, Phytoalexins in defense against pathogens. *Trends Plant Sci* **17**, 73 (Feb, 2012).
19. R. Schuhegger *et al.*, CYP71B15 (PAD3) catalyzes the final step in camalexin biosynthesis. *Plant Physiol* **141**, 1248 (Aug, 2006).

20. R. Mezencev, T. Updegrove, P. Kutschy, M. Repovská, J. F. McDonald, Camalexin induces apoptosis in T-leukemia Jurkat cells by increased concentration of reactive oxygen species and activation of caspase-8 and caspase-9. *Journal of Natural Medicines*, (2011).
21. D. B. Clarke, Glucosinolates, structures and analysis in food. *Analytical Methods* **2**, 310 (2010).
22. U. Wittstock, M. Burow, in *The Arabidopsis book*. (The American Society of Plant Biologists, 2010), vol. 8.
23. E. Andreasson, L. B. Jørgensen, A.-S. Höglund, L. Rask, J. Meijer, Different Myrosinase and Idioblast Distribution in Arabidopsis and Brassica napus1. *PLANT PHYSIOLOGY* **127**, 1750 (2001).
24. H. Husebye, S. Chadchawan, P. Winge, O. P. Thangstad, A. M. Bones, Guard cell- and phloem idioblast-specific expression of thioglucoside glucohydrolase 1 (myrosinase) in Arabidopsis. *Plant Physiol* **128**, 1180 (Apr, 2002).
25. A. M. Bones, J. T. Rossiter, The enzymic and chemically induced decomposition of glucosinolates. *Phytochemistry* **67**, 1053 (Jun, 2006).
26. J. D. Hayes, M. O. Kelleher, I. M. Eggleston, The cancer chemopreventive actions of phytochemicals derived from glucosinolates. *European Journal of Nutrition* **47**, 73 (2008).
27. U. Wittstock, M. Burow, Tipping the scales--specifier proteins in glucosinolate hydrolysis. *IUBMB Life* **59**, 744 (Dec, 2007).
28. B. Lüthy, P. Matile, The mustard oil bomb: rectified analyssis of the subcellular organization of the myrosinase system. *Biochemie und Physiologie der Pflanzen* **179**, 5 (1984).
29. P. A. Blau, P. Feeny, L. Contardo, D. S. Robson, Allylglucosinolate and herbivorous caterpillars: a contrast in toxicity and tolerance. *Science* **200**, 1296 (Jun 16, 1978).
30. K. F. Tierens *et al.*, Study of the role of antimicrobial glucosinolate-derived isothiocyanates in resistance of Arabidopsis to microbial pathogens. *Plant Physiol* **125**, 1688 (Apr, 2001).
31. L. Rask *et al.*, Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant Mol Biol* **42**, 93 (Jan, 2000).
32. G. Brader, E.-. Tas, E. T. Palva, Jasmonate-Dependent Induction of Indole Glucosinolates in Arabidopsis by Culture Filtrates of the Nonspecific Pathogen Erwinia carotovora1. *PLANT PHYSIOLOGY* **126**, 849 (2001).
33. K. Schlaeppi, E. Abou-Mansour, A. Buchala, F. Mauch, Disease resistance of Arabidopsis to Phytophthora brassicae is established by the sequential action of indole glucosinolates and camalexin. *Plant J* **62**, 840 (Jun 1, 2010).
34. S. Buskov, B. Serra, E. Rosa, H. Sørensen, J. C. Sørensen, Effects of Intact Glucosinolates and Products Produced from Glucosinolates in Myrosinase-Catalyzed Hydrolysis on the Potato Cyst Nematode (Globodera rostochiensisCv. Woll). *Journal of Agricultural and Food Chemistry* **50**, 690 (2002).
35. L. Lazzeri, O. Leoni, L. M. Manici, Biocidal plant dried pellets for biofumigation. *Industrial Crops and Products* **20**, 59 (2004).
36. P. Bednarek *et al.*, A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science* **323**, 101 (Jan 2, 2009).

37. N. K. Clay, A. M. Adio, C. Denoux, G. Jander, F. M. Ausubel, Glucosinolate metabolites required for an Arabidopsis innate immune response. *Science* **323**, 95 (Jan 2, 2009).
38. B. A. Halkier, J. Gershenzon, Biology and biochemistry of glucosinolates. *Annu Rev Plant Biol* **57**, 303 (2006).
39. C. D. Grubb, S. Abel, Glucosinolate metabolism and its control. *Trends in Plant Science* **11**, 89 (2006).
40. U. Wittstock, B. A. Halkier, Glucosinolate research in the Arabidopsis era. *Trends Plant Sci* **7**, 263 (Jun, 2002).
41. P. Talalay, J. W. Fahey, Phytochemicals from cruciferous plants protect against cancer by modulating carcinogen metabolism. *J Nutr* **131**, 3027S (Nov, 2001).
42. S. L. Navarro, F. Li, J. W. Lampe, Mechanisms of action of isothiocyanates in cancer chemoprevention: an update. *Food Funct* **2**, 579 (Oct, 2011).
43. Y. S. Keum, W. S. Jeong, A. N. T. Kong, Chemoprevention by isothiocyanates and their underlying molecular signaling mechanisms. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* **555**, 191 (Nov 2, 2004).
44. J. W. Fahey, Y. Zhang, P. Talalay, Broccoli sprouts: an exceptionally rich source of inducers of enzymes that protect against chemical carcinogens. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 10367 (1997).
45. C. Scholl, B. D. Eshelman, D. M. Barnes, P. R. Hanlon, Raphasatin Is a More Potent Inducer of the Detoxification Enzymes Than Its Degradation Products. *Journal of Food Science* **76**, C504 (Apr, 2011).
46. L. Tang *et al.*, Intake of cruciferous vegetables modifies bladder cancer survival. *Cancer Epidemiol Biomarkers Prev* **19**, 1806 (Jul, 2010).
47. D. S. Michaud *et al.*, Fruit and Vegetable Intake and Incidence of Bladder Cancer in a Male Prospective Cohort. *Journal of the National Cancer Institute* **91**, (1999).
48. J. H. Cohen, A. R. Kristal, J. L. Stanford, Fruit and vegetable intakes and prostate cancer risk. *J Natl Cancer Inst* **92**, 61 (Jan 5, 2000).
49. Y. Zhang, P. Talalay, C. G. Cho, G. H. Posner, A major inducer of anticarcinogenic protective enzymes from broccoli: isolation and elucidation of structure. *Proc Natl Acad Sci U S A* **89**, 2399 (Mar 15, 1992).
50. Y. Zhang, The molecular basis that unifies the metabolism, cellular uptake and chemopreventive activities of dietary isothiocyanates. *Carcinogenesis* **33**, 2 (Jan, 2012).
51. Y. Ding *et al.*, Sulforaphane inhibits 4-aminobiphenyl-induced DNA damage in bladder cells and tissues. *Carcinogenesis* **31**, 1999 (Nov, 2010).
52. J. W. Fahey, P. Talalay, Antioxidant functions of sulforaphane: a potent inducer of phase II detoxification enzymes. *Food and Chemical Toxicology* **37**, 973 (1999).
53. A. F. Abdull Razis, M. Bagatta, G. R. De Nicola, R. Iori, C. Ioannides, Intact glucosinolates modulate hepatic cytochrome P450 and phase II conjugation activities and may contribute directly to the chemopreventive activity of cruciferous vegetables. *Toxicology* **277**, 74 (Nov 9, 2010).

54. M. O. Kelleher *et al.*, 1-Cyano-2,3-epithiopropane is a novel plant-derived chemopreventive agent which induces cytoprotective genes that afford resistance against the genotoxic alpha,beta-unsaturated aldehyde acrolein. *Carcinogenesis* **30**, 1754 (Oct, 2009).
55. M. A. Wallig, K. M. Heinz-Taheny, D. L. Epps, T. Gossman, Synergy among Phytochemicals within Crucifers: Does It Translate into Chemoprotection? *journal of nutrition* **135**, 2972S (2005).
56. Y. S. Kim, J. A. Milner, Targets for indole-3-carbinol in cancer prevention. *J Nutr Biochem* **16**, 65 (Feb, 2005).
57. C. W. Nho, E. Jeffery, Crambene, a bioactive nitrile derived from glucosinolate hydrolysis, acts via the antioxidant response element to up-regulate quinone reductase alone or synergistically with indole-3-carbinol. *Toxicol Appl Pharmacol* **198**, 40 (Jul 1, 2004).
58. C. W. Nho, E. Jeffery, The synergistic upregulation of phase II detoxification enzymes by glucosinolate breakdown products in cruciferous vegetables. *Toxicol Appl Pharmacol* **174**, 146 (Jul 15, 2001).
59. A. S. Keck, R. Staack, E. H. Jeffery, The cruciferous nitrile crambene has bioactivity similar to sulforaphane when administered to Fischer 344 rats but is far less potent in cell culture. *Nutr Cancer* **42**, 233 (2002).
60. N. Agerbirk, M. Vos, J. H. Kim, G. Jander, Indole glucosinolate breakdown and its biological effects. *Phytochemistry Reviews* **8**, 101 (2008).
61. B. B. Aggarwal, H. Ichikawa, Molecular targets and anticancer potential of indole-3-carbinol and its derivatives. *cell cycle* **4**, 1201 (Sep, 2005).
62. F. H. Sarkar, Y. Li, Indole-3-carbinol and prostate cancer. *J Nutr* **134**, 3493S (Dec, 2004).
63. S. Banerjee *et al.*, Attenuation of multi-targeted proliferation-linked signaling by 3,3'-diindolylmethane (DIM): from bench to clinic. *Mutat Res* **728**, 47 (Jul-Oct, 2011).
64. T. Vogt, Phenylpropanoid biosynthesis. *Mol Plant* **3**, 2 (Jan, 2010).
65. K. Kai, B. Shimizu, M. Mizutani, K. Watanabe, K. Sakata, Accumulation of coumarins in Arabidopsis thaliana. *Phytochemistry* **67**, 379 (Feb, 2006).
66. M. Adrian, P. Jeandet, J. Veneau, L. A. Weston, R. Bessis, Biological Activity of Resveratrol, a Stilbenic Compound from Grapevines, Against Botrytis cinerea, the Causal Agent for Gray Mold. *Journal of Chemical Ecology* **23**, 1689 (1997).
67. M. Jang, Cancer Chemopreventive Activity of Resveratrol, a Natural Product Derived from Grapes. *Science* **275**, 218 (1997).
68. A. Bishayee, Cancer prevention and treatment with resveratrol: from rodent studies to clinical trials. *Cancer Prev Res (Phila)* **2**, 409 (May, 2009).
69. J. M. Smoliga, J. A. Baur, H. A. Hausenblas, Resveratrol and health--a comprehensive review of human clinical trials. *Mol Nutr Food Res* **55**, 1129 (Aug, 2011).
70. H. H. Chow *et al.*, Resveratrol modulates drug- and carcinogen-metabolizing enzymes in a healthy volunteer study. *Cancer Prev Res (Phila)* **3**, 1168 (Sep, 2010).
71. M. R. Grant, J. D. Jones, Hormone (dis)harmony moulds plant health and disease. *Science* **324**, 750 (May 8, 2009).

72. M. C. Wildermuth, Variations on a theme: synthesis and modification of plant benzoic acids. *Curr Opin Plant Biol* **9**, 288 (Jun, 2006).
73. M. Reichelt *et al.*, Benzoic acid glucosinolate esters and other glucosinolates from Arabidopsis thaliana. *Phytochemistry* **59**, 663 (2002).
74. G. G. Duthie, A. D. Wood, Natural salicylates: foods, functions and disease prevention. *Food Funct* **2**, 515 (Sep, 2011).
75. J. K. Kundu, Y. J. Surh, Nrf2-Keap1 Signaling as a Potential Target for Chemoprevention of Inflammation-Associated Carcinogenesis. *Pharmaceutical Research* **27**, 999 (Jun, 2010).
76. O. Fingrut, E. Flescher, Plant stress hormones suppress the proliferation and induce apoptosis in human cancer cells. *Leukemia* **16**, 608 (Apr, 2002).
77. G. R. Beecher, Overview of dietary flavonoids: nomenclature, occurrence and intake. *J Nutr* **133**, 3248S (Oct, 2003).
78. C. S. Buer, N. Imin, M. A. Djordjevic, Flavonoids: New Roles for Old Molecules. *Journal of Integrative Plant Biology* **52**, 98 (2010).
79. I. Hernández, F. Van Breusegem, Opinion on the possible role of flavonoids as energy escape valves: Novel tools for nature's Swiss army knife? *Plant Science* **179**, 297 (2010).
80. M. Slana, D. Zigon, T. Makovec, H. Lenasi, The response of filamentous fungus Rhizopus nigricans to flavonoids. *J Basic Microbiol* **51**, 433 (Aug, 2011).
81. L. J. Shaw, P. Morris, J. E. Hooker, Perception and modification of plant flavonoid signals by rhizosphere microorganisms. *Environ Microbiol* **8**, 1867 (Nov, 2006).
82. D. Treutter, Significance of flavonoids in plant resistance and enhancement of their biosynthesis. *Plant Biol (Stuttg)* **7**, 581 (Nov, 2005).
83. J. Li, T. M. Ou-Lee, R. Raba, R. G. Amundson, R. L. Last, Arabidopsis Flavonoid Mutants Are Hypersensitive to UV-B Irradiation. *Plant Cell* **5**, 171 (Feb, 1993).
84. M. Jacobs, P. H. Rubery, Naturally occurring auxin transport regulators. *Science* **241**, 346 (Jul 15, 1988).
85. D. E. Brown *et al.*, Flavonoids act as negative regulators of auxin transport in vivo in Arabidopsis. *PLANT PHYSIOLOGY* **126**, 524 (Jun, 2001).
86. B. Winkel-Shirley, Biosynthesis of flavonoids and effects of stress. *Curr Opin Plant Biol* **5**, 218 (Jun, 2002).
87. C. S. Buer, N. Imin, M. A. Djordjevic, Flavonoids: new roles for old molecules. *J Integr Plant Biol* **52**, 98 (Jan, 2010).
88. K. M. Lee *et al.*, Kaempferol inhibits UVB-induced COX-2 expression by suppressing Src kinase activity. *Biochem Pharmacol* **80**, 2042 (Dec 15, 2010).
89. K. M. Lee *et al.*, Phosphatidylinositol 3-kinase, a novel target molecule for the inhibitory effects of kaempferol on neoplastic cell transformation. *Carcinogenesis* **31**, 1338 (Aug, 2010).
90. Y. Y. Cho *et al.*, A regulatory mechanism for RSK2 NH(2)-terminal kinase activity. *Cancer Res* **69**, 4398 (May 15, 2009).
91. K. W. Lee *et al.*, Raf and MEK protein kinases are direct molecular targets for the chemopreventive effect of quercetin, a major flavonol in red wine. *Cancer Res* **68**, 946 (Feb 1, 2008).

92. M. K. Hwang, N. R. Song, N. J. Kang, K. W. Lee, H. J. Lee, Activation of phosphatidylinositol 3-kinase is required for tumor necrosis factor-alpha-induced upregulation of matrix metalloproteinase-9: its direct inhibition by quercetin. *Int J Biochem Cell Biol* **41**, 1592 (Jul, 2009).
93. K. M. Lee, M. K. Hwang, D. E. Lee, K. W. Lee, H. J. Lee, Protective effect of quercetin against arsenite-induced COX-2 expression by targeting PI3K in rat liver epithelial cells. *J Agric Food Chem* **58**, 5815 (May 12, 2010).
94. A. J. Vargas, R. Burd, Hormesis and synergy: pathways and mechanisms of quercetin in cancer prevention and management. *Nutr Rev* **68**, 418 (Jul, 2010).
95. D. Tholl, S. Lee, Terpene Specialized Metabolism in Arabidopsis thaliana. *Arabidopsis Book* **9**, e0143 (2011).
96. M. Hasegawa *et al.*, Phytoalexin accumulation in the interaction between rice and the blast fungus. *Mol Plant Microbe Interact* **23**, 1000 (Aug, 2010).
97. C. I. Keeling, J. Bohlmann, Diterpene resin acids in conifers. *Phytochemistry* **67**, 2415 (Nov, 2006).
98. S. Heiling *et al.*, Jasmonate and ppHsystemin regulate key Malonylation steps in the biosynthesis of 17-Hydroxygeranyllinalool Diterpene Glycosides, an abundant and effective direct defense against herbivores in Nicotiana attenuata. *Plant Cell* **22**, 273 (Jan, 2010).
99. T. C. Turlings, J. Ton, Exploiting scents of distress: the prospect of manipulating herbivore-induced plant odours to enhance the control of agricultural pests. *Curr Opin Plant Biol* **9**, 421 (Aug, 2006).
100. S. B. Unsicker, G. Kunert, J. Gershenzon, Protective perfumes: the role of vegetative volatiles in plant defense against herbivores. *Curr Opin Plant Biol* **12**, 479 (Aug, 2009).
101. D. Tholl, F. Chen, J. Petri, J. Gershenzon, E. Pichersky, Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. *Plant J* **42**, 757 (Jun, 2005).
102. F. Chen *et al.*, Biosynthesis and emission of terpenoid volatiles from Arabidopsis flowers. *Plant Cell* **15**, 481 (Feb, 2003).
103. F. Loreto, P. Pinelli, F. Manes, H. Kollist, Impact of ozone on monoterpene emissions and evidence for an isoprene-like antioxidant action of monoterpenes emitted by Quercus ilex leaves. *Tree Physiology* **24**, 361 (2004).
104. A. V. Rao, L. G. Rao, Carotenoids and human health. *Pharmacol Res* **55**, 207 (Mar, 2007).
105. J. Shah, Lipids, lipases, and lipid-modifying enzymes in plant disease resistance. *Annu Rev Phytopathol* **43**, 229 (2005).
106. R. M. Van Poecke, M. A. Posthumus, M. Dicke, Herbivore-induced volatile production by Arabidopsis thaliana leads to attraction of the parasitoid Cotesia rubecula: chemical, behavioral, and gene-expression analysis. *J Chem Ecol* **27**, 1911 (Oct, 2001).
107. R. G. Upchurch, Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. *Biotechnol Lett* **30**, 967 (Jun, 2008).
108. A. Raffaele, A. Leger, D. Roby, Very long chain fatty acid and lipid signaling in the response of plants to pathogens. *Plant Signaling & Behavior* **4**, 94 (2009).

109. H. Vesper *et al.*, Sphingolipids in food and the emerging importance of sphingolipids to nutrition. *J Nutr* **129**, 1239 (Jul, 1999).
110. C. Chen, A. N. Kong, Dietary chemopreventive compounds and ARE/EpRE signaling. *Free Radic Biol Med* **36**, 1505 (Jun 15, 2004).
111. M. K. Kwak, T. W. Kensler, Targeting NRF2 signaling for cancer chemoprevention. *Toxicol Appl Pharmacol* **244**, 66 (Apr 1, 2010).
112. M. Zhu, W. E. Fahl, Development of a green fluorescent protein microplate assay for the screening of chemopreventive agents. *Anal Biochem* **287**, 210 (Dec 15, 2000).
113. A. M. Boerboom *et al.*, Newly constructed stable reporter cell lines for mechanistic studies on electrophile-responsive element-mediated gene expression reveal a role for flavonoid planarity. *Biochem Pharmacol* **72**, 217 (Jul 14, 2006).
114. Gross, Dalebout, Grubb, Abel, Functional detection of chemopreventive glucosinolates in Arabidopsis thaliana. *Plant Science (Shannon, Ireland)* **159**, 265 (2000).
115. C. Grubb, H. Gross, D. Chen, S. Abel, Identification of Arabidopsis mutants with altered glucosinolate profiles based on isothiocyanate bioactivity. *Plant Science* **162**, 143 (2002).
116. Q. Wang, C. D. Grubb, S. Abel, Direct analysis of single leaf disks for chemopreventive glucosinolates. *Phytochem Anal* **13**, 152 (May-Jun, 2002).
117. H. J. Prochaska, A. B. Santamaria, Direct measurement of NAD(P)H:quinone reductase from cells cultured in microtiter wells: a screening assay for anticarcinogenic enzyme inducers. *Anal Biochem* **169**, 328 (Mar, 1988).
118. H. J. Prochaska, A. B. Santamaria, P. Talalay, Rapid detection of inducers of enzymes that protect against carcinogens. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 2394 (1992).
119. X. Feng, X. Liu, Q. Luo, B. F. Liu, Mass spectrometry in systems biology: an overview. *Mass Spectrom Rev* **27**, 635 (Nov-Dec, 2008).
120. Y. Okazaki, K. Saito, Recent advances of metabolomics in plant biotechnology. *Plant Biotechnol Rep* **6**, 1 (Jan, 2012).
121. S. G. Villas-Boas, S. Mas, M. Akesson, J. Smedsgaard, J. Nielsen, Mass spectrometry in metabolome analysis. *Mass Spectrom Rev* **24**, 613 (Sep-Oct, 2005).
122. W. B. Dunn, Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys Biol* **5**, 011001 (2008).
123. R. A. Dixon *et al.*, Applications of metabolomics in agriculture. *J Agric Food Chem* **54**, 8984 (Nov 29, 2006).
124. R. C. H. de Vos *et al.*, Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *nature protocols* **2**, 778 (2007).
125. S. Moco *et al.*, A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol* **141**, 1205 (Aug, 2006).
126. V. V. Tolstikov, A. Lommen, K. Nakanishi, N. Tanaka, O. Fiehn, Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. *Anal Chem* **75**, 6737 (Dec 1, 2003).

127. H. Rischer *et al.*, Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in Catharanthus roseus cells. *Proc Natl Acad Sci U S A* **103**, 5614 (Apr 4, 2006).
128. D. V. Huhman, L. W. Sumner, Metabolic profiling of saponins in Medicago sativa and Medicago truncatula using HPLC coupled to an electrospray ion-trap mass spectrometer. *Phytochemistry* **59**, 347 (2002).
129. C. Bottcher *et al.*, Metabolome analysis of biosynthetic mutants reveals a diversity of metabolic changes and allows identification of a large number of new compounds in Arabidopsis. *Plant Physiol* **147**, 2107 (Aug, 2008).
130. J. T. Odell, F. Nagy, N.-H. Chua, Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* **313**, 810 (1985).
131. C. Bolle, A. Schneider, D. Leister, Perspectives on Systematic Analyses of Gene Function in Arabidopsis thaliana: New Tools, Topics and Trends. *Curr Genomics* **12**, 1 (Mar, 2011).
132. N. Marsch-Martinez, A transposon-based activation tagging system for gene function discovery in Arabidopsis. *Methods Mol Biol* **754**, 67 (2011).
133. Walden R *et al.*, Activation tagging: a means of isolating genes implicated as playing a role in plant growth and development. *Plant Molecular Biology* **26**, 1521 (1994).
134. M. Nakazawa *et al.*, Activation tagging, a novel tool to dissect the functions of a gene family. *Plant J* **34**, 741 (Jun, 2003).
135. A. Schneider *et al.*, A transposon-based activation-tagging population in Arabidopsis thaliana (TAMARA) and its application in the identification of dominant developmental and metabolic mutations. *FEBS Lett* **579**, 4622 (Aug 29, 2005).
136. D. Weigel *et al.*, Activation tagging in Arabidopsis. *Plant Physiol* **122**, 1003 (Apr, 2000).
137. A. F. Tissier *et al.*, Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: a tool for functional genomics. *Plant Cell* **11**, 1841 (Oct, 1999).
138. N. Marsch-Martinez, R. Greco, G. Van Arkel, L. Herrera-Estrella, A. Pereira, Activation tagging using the En-I maize transposon system in Arabidopsis. *Plant Physiol* **129**, 1544 (Aug, 2002).
139. K. Wilson, D. Long, J. Swinburne, G. Coupland, A Dissociation insertion causes a semidominant mutation that increases expression of TINY, an Arabidopsis gene related to APETALA2. *Plant Cell* **8**, 659 (Apr, 1996).
140. S. J. Robinson *et al.*, An archived activation tagged population of Arabidopsis thaliana to facilitate forward genetics approaches. *BMC Plant Biol* **9**, 101 (2009).
141. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**, 796 (Dec 14, 2000).
142. R. J. Bino *et al.*, Potential of metabolomics as a functional genomics tool. *Trends in Plant Science* **9**, 418 (2004).
143. M. M. Neff, Inaugural Article: BAS1: A gene regulating brassinosteroid levels and light responsiveness in Arabidopsis. *Proceedings of the National Academy of Sciences* **96**, 15316 (1999).

144. E. van der Graaff, A. D. Dulk-Ras, P. J. Hooykaas, B. Keller, Activation tagging of the LEAFY PETIOLE gene affects leaf petiole development in Arabidopsis thaliana. *Development* **127**, 4971 (Nov, 2000).
145. M. De Block *et al.*, Engineering herbicide resistance in plants by expression of a detoxifying enzyme. *EMBO Journal* **6**, 2513 (1987).
146. D. P. O'Keefe *et al.*, Plant Expression of a Bacterial Cytochrome P450 That Catalyzes Activation of a Sulfonylurea Pro-Herbicide. *PLANT PHYSIOLOGY* **105**, 473 (1994).
147. T. Nakagawa *et al.*, Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *Journal of Bioscience and Bioengineering* **104**, 34 (Jul, 2007).
148. B. Berger, Stracke, R., Yatusevich, R., Weisshaar, B., Flugge, U. I. and Gigolashvili, T., A simplified method for the analysis of transcription factor-promoter interactions that allows highthroughput data generation. *The Plant Journal* **50**, 911 (2007).
149. R. Stracke *et al.*, Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the Arabidopsis thaliana seedling. *Plant J* **50**, 660 (May, 2007).
150. M. Todesco, I. Rubio-Somoza, J. Paz-Ares, D. Weigel, A collection of target mimics for comprehensive analysis of microRNA function in Arabidopsis thaliana. *PLoS Genet* **6**, e1001031 (Jul, 2010).
151. T. Gigolashvili *et al.*, The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in Arabidopsis thaliana. *Plant J* **50**, 886 (Jun, 2007).
152. D. Hanahan, Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* **166**, 557 (Jun 5, 1983).
153. C. Koncz, J. Schell, The Promoter of Tl-DNA Gene 5 Controls the Tissue-Specific Expression of Chimeric Genes Carried by a Novel Type of Agrobacterium Binary Vector. *Molecular & General Genetics* **204**, 383 (1986).
154. O. Voinnet, S. Rivas, P. Mestre, D. Baulcombe, An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J* **33**, 949 (Mar, 2003).
155. L. van der Fits, Deakin, E. A., Hoge, J. H. C. and Memelink, J. , The ternary transformation system: constitutive virG on a compatible plasmid dramatically increases Agrobacteriummediated plant transformation. *Plant Molecular Biology* **43**, 495 (2000).
156. R. D. Sambrook J, *Molecular cloning: a laboratory manual.* . (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2001).
157. K. Edwards *et al*, A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res* **19**, 1349 (Mar 25, 1991).
158. P. Lamesch *et al.*, The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res* **40**, D1202 (Jan, 2012).
159. R. Schwacke *et al.*, ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. *Plant Physiol* **131**, 16 (Jan, 2003).
160. R. K. Saiki *et al.*, Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487 (Jan 29, 1988).

161. R. K. Saiki *et al.*, Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle-Cell Anemia. *Science* **230**, 1350 (1985).
162. C. Varotto *et al.*, Disruption of the Arabidopsis photosystem I gene psaE1 affects photosynthesis and impairs growth. *Plant J* **22**, 115 (Apr, 2000).
163. O. A. Koroleva, M. L. Tomlinson, D. Leader, P. Shaw, J. H. Doonan, High-throughput protein localization in Arabidopsis using Agrobacterium-mediated transient expression of GFP-ORF fusions. *Plant journal* **41**, 162 (Jan, 2005).
164. E. von Roepenack-Lahaye *et al.*, Profiling of Arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. *Plant Physiol* **134**, 548 (Feb, 2004).
165. S. Krizowski, Universität zu Köln (2008).
166. E. G. Rogan, The natural chemopreventive compound indole-3-carbinol: state of the science. *in vivo* **20**, 221 (Mar-Apr, 2006).
167. C. Eigen, University of Cologne (2010).
168. B. T. S. Glynn Dennis Jr., Douglas A. Hosack, Jun Yang, Michael W. Baseler, H. Clifford Lane, Richard A. Lempicki, DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology* **4**, 3 (2003).
169. S. Disch *et al.*, The E3 ubiquitin ligase BIG BROTHER controls arabidopsis organ size in a dosage-dependent manner. *Curr Biol* **16**, 272 (Feb 7, 2006).
170. L. Marbaise, Universität zu Köln (2010).
171. Z. Xu *et al.*, Functional genomic analysis of Arabidopsis thaliana glycoside hydrolase family 1. *Plant Mol Biol* **55**, 343 (May, 2004).
172. I. Nitz, H. Berkefeld, P. S. Puzio, F. M. Grundler, Pyk10, a seedling and root specific gene and promoter from Arabidopsis thaliana. *Plant Sci* **161**, 337 (Jul, 2001).
173. R. Matsushima, Y. Fukao, M. Nishimura, I. Hara-Nishimura, NAI1 gene encodes a basic-helix-loop-helix-type putative transcription factor that regulates the formation of an endoplasmic reticulum-derived structure, the ER body. *Plant Cell* **16**, 1536 (Jun, 2004).
174. Y. O. Ahn *et al.*, Scopolin-hydrolyzing beta-glucosidases in roots of Arabidopsis. *Plant Cell Physiol* **51**, 132 (Jan, 2010).
175. T. Obayashi, S. Hayashi, M. Saeki, H. Ohta, K. Kinoshita, ATTED-II provides coexpressed gene networks for Arabidopsis. *Nucleic Acids Res* **37**, D987 (Jan, 2009).
176. T. Obayashi, K. Nishida, K. Kasahara, K. Kinoshita, ATTED-II updates: condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants. *Plant Cell Physiol* **52**, 213 (Feb, 2011).
177. M. Pfalz *et al.*, Metabolic engineering in Nicotiana benthamiana reveals key enzyme functions in Arabidopsis indole glucosinolate modification. *Plant Cell* **23**, 716 (Feb, 2011).
178. H. Frerigmann, Universität zu Köln (2011).
179. R. Matsushima, M. Kondo, M. Nishimura, I. Hara-Nishimura, A novel ER-derived compartment, the ER body, selectively accumulates a beta-

glucosidase with an ER-retention signal in Arabidopsis. *Plant J* **33**, 493 (Feb, 2003).
180. A. J. Nagano, Y. Fukao, M. Fujiwara, M. Nishimura, I. Hara-Nishimura, Antagonistic jacalin-related lectins regulate the size of ER body-type beta-glucosidase complexes in Arabidopsis thaliana. *Plant Cell Physiol* **49**, 969 (Jun, 2008).
181. G. Gonnella, Eberhard-Karls-Universität Tuebingen (2004).
182. A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567 (Jan 19, 2001).
183. D. Winter *et al.*, An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* **2**, e718 (2007).
184. T. Masaki *et al.*, Activation tagging of a gene for a protein with novel class of CCT-domain activates expression of a subset of sugar-inducible genes in Arabidopsis thaliana. *Plant journal* **43**, 142 (Jul, 2005).
185. Z. Zhang, M. Wang, Z. Li, Q. Li, Z. He, Arabidopsis GH3.5 regulates salicylic acid-dependent and both NPR1-dependent and independent defense responses. *Plant Signaling & Behavior* **3**, 537 (2008).
186. J. E. Park *et al.*, GH3-mediated auxin homeostasis links growth regulation with stress adaptation response in Arabidopsis. *J Biol Chem* **282**, 10036 (Mar 30, 2007).
187. M. Nakazawa *et al.*, DFL1, an auxin-responsive GH3 gene homologue, negatively regulates shoot cell elongation and lateral root formation, and positively regulates the light response of hypocotyl length. *The Plant Journal: For Cell and Molecular Biology* **25**, 213 (2001).
188. P. E. Staswick *et al.*, Characterization of an Arabidopsis Enzyme Family That Conjugates Amino Acids to Indole-3-Acetic Acid. *THE PLANT CELL ONLINE* **17**, 616 (2005).
189. S. Ossowski, J. Fitz, R. Schwab, M. Riester, D. Weigel, personal communication.
190. C. Addo-Quaye, T. W. Eshoo, D. P. Bartel, M. J. Axtell, Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Curr Biol* **18**, 758 (May 20, 2008).
191. N. Fahlgren *et al.*, High-Throughput Sequencing of Arabidopsis microRNAs: Evidence for Frequent Birth and Death of MIRNA Genes. *PLoS ONE* **2**, e219 (2007).
192. E. Allen *et al.*, Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. *Nat Genet* **36**, 1282 (Dec, 2004).
193. Y. Zhang, W. K. Jiang, L. Z. Gao, Evolution of microRNA genes in Oryza sativa and Arabidopsis thaliana: an update of the inverted duplication model. *PLoS ONE* **6**, e28073 (2011).
194. M. Megraw *et al.*, MicroRNA promoter element discovery in Arabidopsis. *RNA* **12**, 1612 (Sep, 2006).
195. J. Krol, I. Loedige, W. Filipowicz, The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* **11**, 597 (Sep, 2010).
196. R. Stracke *et al.*, The Arabidopsis bZIP transcription factor HY5 regulates expression of the PFG1/MYB12 gene in response to light and ultraviolet-B radiation. *Plant Cell Environ* **33**, 88 (Jan, 2010).

197. H. Zhang *et al.*, Genome-wide mapping of the HY5-mediated genenetworks in Arabidopsis that involve both transcriptional and post-transcriptional regulation. *The Plant Journal* **65**, 346 (2011).
198. X. L. Tan, M. Shi, H. Tang, W. Han, S. D. Spivack, Candidate dietary phytochemicals modulate expression of phase II enzymes GSTP1 and NQO1 in human lung cells. *J Nutr* **140**, 1404 (Aug, 2010).
199. L. L. Song *et al.*, Cancer chemopreventive activity mediated by 4'-bromoflavone, a potent inducer of phase II detoxification enzymes. *Cancer Res* **59**, 578 (Feb 1, 1999).
200. M. R. Saracino, J. W. Lampe, Phytochemical regulation of UDP-glucuronosyltransferases: implications for cancer prevention. *Nutr Cancer* **59**, 121 (2007).
201. Y. Y. Lee-Hilz *et al.*, Pro-oxidant activity of flavonoids induces EpRE-mediated gene expression. *Chem Res Toxicol* **19**, 1499 (Nov, 2006).
202. S. T. van Cruchten *et al.*, The role of epoxidation and electrophile-responsive element-regulated gene transcription in the potentially beneficial and harmful effects of the coffee components cafestol and kahweol. *J Nutr Biochem* **21**, 757 (Aug, 2010).
203. W. Brand *et al.*, Stereoselective conjugation, transport and bioactivity of s- and R-hesperetin enantiomers in vitro. *J Agric Food Chem* **58**, 6119 (May 26, 2010).
204. M. Vermeulen *et al.*, Potency of isothiocyanates to induce luciferase reporter gene expression via the electrophile-responsive element from murine glutathione S-transferase Ya. *Toxicol In Vitro* **23**, 617 (Jun, 2009).
205. M. Muzolf-Panek *et al.*, Role of catechin quinones in the induction of EpRE-mediated gene expression. *Chem Res Toxicol* **21**, 2352 (Dec, 2008).
206. V. A. van Beelen *et al.*, Differential induction of electrophile-responsive element-regulated genes by n-3 and n-6 polyunsaturated fatty acids. *FEBS Lett* **580**, 4587 (Aug 21, 2006).
207. Y. Y. Lee-Hilz, M. Stolaki, W. J. van Berkel, J. M. Aarts, I. M. Rietjens, Activation of EpRE-mediated gene transcription by quercetin glucuronides depends on their deconjugation. *Food Chem Toxicol* **46**, 2128 (Jun, 2008).
208. J. Yang, R. H. Liu, Induction of phase II enzyme, quinone reductase, in murine hepatoma cells in vitro by grape extracts and selected phytochemicals. *Food Chemistry* **114**, 898 (2009).
209. M. Prince *et al.*, Comparison of citrus coumarins on carcinogen-detoxifying enzymes in Nrf2 knockout mice. *Toxicol Lett* **185**, 180 (Mar 28, 2009).
210. R. H. Liu, Potential synergy of phytochemicals in cancer prevention: mechanism of action. *J Nutr* **134**, 3479S (Dec, 2004).
211. M. Jahangir, I. B. Abdel-Farid, H. K. Kim, Y. H. Choi, R. Verpoorte, Healthy and unhealthy plants: The effect of stress on the metabolism of Brassicaceae. *Environmental and Experimental Botany* **67**, 23 (2009).
212. A. Gopalakrishnan, A. N. Tony Kong, Anticarcinogenesis by dietary phytochemicals: cytoprotection by Nrf2 in normal cells and cytotoxicity by modulation of transcription factors NF-kappa B and AP-1 in abnormal cancer cells. *Food Chem Toxicol* **46**, 1257 (Apr, 2008).
213. M. H. Traka, R. F. Mithen, Plant science and human nutrition: challenges in assessing health-promoting properties of phytochemicals. *Plant Cell* **23**, 2483 (Jul, 2011).

214. G. M. DeNicola *et al.*, Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. *Nature* **475**, 106 (2011).
215. Z. Minic, Physiological roles of plant glycoside hydrolases. *Planta* **227**, 723 (Mar, 2008).
216. R. Aharon, Overexpression of a Plasma Membrane Aquaporin in Transgenic Tobacco Improves Plant Vigor under Favorable Growth Conditions but Not under Drought or Salt Stress. *THE PLANT CELL ONLINE* **15**, 439 (2003).
217. B. Schulz, Functional Classification of Plant Plasma Membrane Transporters. **19**, 131 (2011).
218. K. Yazaki, Transporters of secondary metabolites. *Curr Opin Plant Biol* **8**, 301 (Jun, 2005).
219. K. Yazaki, ABC transporters involved in the transport of plant secondary metabolites. *FEBS Lett* **580**, 1183 (Feb 13, 2006).
220. E. Grotewold, Transcription factors for predictive plant metabolic engineering: are we there yet? *Curr Opin Biotechnol* **19**, 138 (Apr, 2008).
221. C. F. Mouchel, G. C. Briggs, C. S. Hardtke, Natural genetic variation in Arabidopsis identifies BREVIS RADIX, a novel regulator of cell proliferation and elongation in the root. *Genes Dev* **18**, 700 (Mar 15, 2004).
222. C. F. Mouchel, K. S. Osmont, C. S. Hardtke, BRX mediates feedback between brassinosteroid levels and auxin signalling in root growth. *Nature* **443**, 458 (Sep 28, 2006).
223. J. Li *et al.*, BREVIS RADIX is involved in cytokinin-mediated inhibition of lateral root initiation in Arabidopsis. *Planta* **229**, 593 (Feb, 2009).
224. L. Kienow *et al.*, Jasmonates meet fatty acids: functional analysis of a new acyl-coenzyme A synthetase family from Arabidopsis thaliana. *J Exp Bot* **59**, 403 (2008).
225. D. P. Dixon, A. Lapthorn, R. Edwards, Plant glutathione transferases. *Genome Biol* **3**, REVIEWS3004 (2002).
226. M. Boudsocq, H. Barbier-Brygoo, C. Lauriere, Identification of nine sucrose nonfermenting 1-related protein kinases 2 activated by hyperosmotic and saline stresses in Arabidopsis thaliana. *J Biol Chem* **279**, 41758 (Oct 1, 2004).
227. J. Zhao, L. C. Davis, R. Verpoorte, Elicitor signal transduction leading to production of plant secondary metabolites. *Biotechnol Adv* **23**, 283 (Jun, 2005).
228. A. El-Kereamy *et al.*, Exogenous ethylene stimulates the long-term expression of genes related to anthocyanin biosynthesis in grape berries. *Physiologia Plantarum* **119**, 175 (2003).
229. L. Onate-Sanchez, J. P. Anderson, J. Young, K. B. Singh, AtERF14, a member of the ERF family of transcription factors, plays a nonredundant role in plant defense. *Plant Physiol* **143**, 400 (Jan, 2007).
230. B. Zambelli, F. Musiani, S. Ciurli, in *Met Ions Life Sci*. (2012), vol. 10, pp. 135-70.
231. E. Lechner, P. Achard, A. Vansiri, T. Potuschak, P. Genschik, F-box proteins everywhere. *Curr Opin Plant Biol* **9**, 631 (Dec, 2006).
232. G. Parry, M. Estelle, Auxin receptors: a new role for F-box proteins. *Curr Opin Cell Biol* **18**, 152 (Apr, 2006).
233. D. E. Somers, Clock-associated genes in Arabidopsis: a family affair. *Philos Trans R Soc Lond B Biol Sci* **356**, 1745 (Nov 29, 2001).

234. K. Ogasawara *et al.*, Constitutive and inducible ER bodies of Arabidopsis thaliana accumulate distinct beta-glucosidases. *Plant Cell Physiol* **50**, 480 (Mar, 2009).
235. I. Hara-Nishimura, R. Matsushima, A wound-inducible organelle derived from endoplasmic reticulum: a plant strategy against environmental stresses? *Curr Opin Plant Biol* **6**, 583 (Dec, 2003).
236. R. Matsushima *et al.*, The ER body, a novel endoplasmic reticulum-derived structure in Arabidopsis. *Plant Cell Physiol* **44**, 661 (Jul, 2003).
237. K. Yamada, A. J. Nagano, K. Ogasawara, I. Hara-Nishimura, M. Nishimura, The ER body, a new organelle in Arabidopsis thaliana, requires NAI2 for its formation and accumulates specific beta-glucosidases. *Plant Signal Behav* **4**, 849 (Sep, 2009).
238. R. Matsushima *et al.*, An endoplasmic reticulum-derived structure that is induced under stress conditions in Arabidopsis. *Plant Physiol* **130**, 1807 (Dec, 2002).
239. A. J. Nagano, R. Matsushima, I. Hara-Nishimura, Activation of an ER-body-localized beta-glucosidase via a cytosolic binding partner in damaged tissues of Arabidopsis thaliana. *Plant Cell Physiol* **46**, 1140 (Jul, 2005).
240. P. Rose, 7-Methylsulfinylheptyl and 8-methylsulfinyloctyl isothiocyanates from watercress are potent inducers of phase II enzymes. *Carcinogenesis* **21**, 1983 (2000).
241. E. S. Hwang, E. H. Jeffery, Induction of quinone reductase by sulforaphane and sulforaphane N-acetylcysteine conjugate in murine hepatoma cells. *J Med Food* **8**, 198 (Summer, 2005).
242. Y. Uda, K. R. Price, G. Williamson, M. J. C. Rhodes, Induction of the anticarcinogenic marker enzyme, quinone reductase, in murine hepatoma cells in vitro by flavonoids. *Cancer Letters* **120**, 213 (1997).
243. A. Schmidt, Friedrich-Schiller-Universität Jena (2007).
244. M. J. Mueller, Archetype signals in plants: the phytoprostanes. *Curr Opin Plant Biol* **7**, 441 (Aug, 2004).
245. C. Loeffler *et al.*, B1-phytoprostanes trigger plant defense and detoxification responses. *Plant Physiol* **137**, 328 (Jan, 2005).
246. M. Heil, Fitness costs of induced resistance: emerging experimental support for a slippery concept. *Trends in Plant Science* **7**, 61 (2002).
247. J. Putterill, F. Robson, K. Lee, R. Simon, G. Coupland, The CONSTANS gene of arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *cell* **80**, 847 (1995).
248. C. Strayer, Cloning of the Arabidopsis Clock Gene TOC1, an Autoregulatory Response Regulator Homolog. *Science* **289**, 768 (2000).
249. N. Nakamichi *et al.*, The Arabidopsis pseudo-response regulators, PRR5 and PRR7, coordinately play essential roles for circadian clock function. *Plant Cell Physiol* **46**, 609 (Apr, 2005).
250. P. A. Salome, J. P. To, J. J. Kieber, C. R. McClung, Arabidopsis response regulators ARR3 and ARR4 play cytokinin-independent roles in the control of circadian period. *Plant Cell* **18**, 55 (Jan, 2006).
251. K. A. Kaczorowski, P. H. Quail, Arabidopsis PSEUDO-RESPONSE REGULATOR7 is a signaling intermediate in phytochrome-regulated seedling deetiolation and phasing of the circadian clock. *Plant Cell* **15**, 2654 (Nov, 2003).

252. O. Zobell, G. Coupland, B. Reiss, The family of CONSTANS-like genes in Physcomitrella patens. *Plant Biol (Stuttg)* **7**, 266 (May, 2005).
253. S. Kurup, H. D. Jones, M. J. Holdsworth, Interactions of the developmental regulator ABI3 with proteins identified from developing Arabidopsis seeds. *The Plant Journal* **21**, 143 (2000).
254. S. Wenkel *et al.*, CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. *Plant Cell* **18**, 2971 (Nov, 2006).
255. J. Ludwig-Muller, Auxin conjugates: their role for plant development and in the evolution of land plants. *Journal of Experimental Botany* **62**, 1757 (2011).
256. P. E. Staswick, I. Tiryaki, M. L. Rowe, Jasmonate response locus JAR1 and several related Arabidopsis genes encode enzymes of the firefly luciferase superfamily that show activity on jasmonic, salicylic, and indole-3-acetic acids in an assay for adenylation. *Plant Cell* **14**, 1405 (Jun, 2002).
257. Z. Zhang *et al.*, Dual regulation role of GH3.5 in salicylic acid and auxin signaling during Arabidopsis-Pseudomonas syringae interaction. *Plant Physiol* **145**, 450 (Oct, 2007).
258. K. Kai, J. Horita, K. Wakasa, H. Miyagawa, Three oxidative metabolites of indole-3-acetic acid from Arabidopsis thaliana. *Phytochemistry* **68**, 1651 (Jun, 2007).
259. R. L. McCarthy, R. Zhong, Z. H. Ye, MYB83 Is a Direct Target of SND1 and Acts Redundantly with MYB46 in the Regulation of Secondary Cell Wall Biosynthesis in Arabidopsis. *Plant and Cell Physiology* **50**, 1950 (2009).
260. C. Dubos *et al.*, MYB transcription factors in Arabidopsis. *Trends in Plant Science* **15**, 573 (2010).
261. V. Kirik, K. Kolle, T. Wohlfarth, S. Misera, H. Baumlein, Ectopic expression of a novel MYB gene modifies the architecture of the Arabidopsis inflorescence. *Plant J* **13**, 729 (Mar, 1998).
262. H. L. Jin, C. Martin, Multifunctionality and diversity within the plant MYB-gene family. *Plant Molecular Biology* **41**, 577 (Nov, 1999).
263. F. Mehrtens, H. Kranz, P. Bednarek, B. Weisshaar, The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol* **138**, 1083 (Jun, 2005).
264. W. L. Kubasek *et al.*, Regulation of Flavonoid Biosynthetic Genes in Germinating Arabidopsis Seedlings. *Plant Cell* **4**, 1229 (Oct, 1992).
265. R. Stracke *et al.*, Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol glycoside accumulation in Arabidopsis thaliana plants reveals MYB11-, MYB12- and MYB111-independent flavonol glycoside accumulation. *New Phytol* **188**, 985 (Dec, 2010).
266. T. Oyama, Y. Shimura, K. Okada, The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. *Genes Dev.* **11**, 2983 (1997).
267. S. Chattopadhyay, L. H. Ang, P. Puente, X. W. Deng, N. Wei, Arabidopsis bZIP protein HY5 directly interacts with light-responsive promoters in mediating light control of gene expression. *Plant Cell* **10**, 673 (May, 1998).

268. M. T. Osterlund, C. S. Hardtke, N. Wei, X. W. Deng, Targeted destabilization of HY5 during light-regulated development of Arabidopsis. *Nature* **405**, 462 (May 25, 2000).
269. R. Ulm *et al.*, Genome-wide analysis of gene expression reveals function of the bZIP transcription factor HY5 in the UV-B response of Arabidopsis. *Proc Natl Acad Sci U S A* **101**, 1397 (Feb 3, 2004).
270. L.-H. Ang *et al.*, Molecular Interaction between COP1 and HY5 Defines a Regulatory Switch for Light Control of Arabidopsis Development. *Molecular Cell* **1**, 213 (1998).
271. A. Oravecz *et al.*, CONSTITUTIVELY PHOTOMORPHOGENIC1 is required for the UV-B response in Arabidopsis. *Plant Cell* **18**, 1975 (Aug, 2006).
272. Y. H. Song *et al.*, DNA-binding study identifies C-box and hybrid C/G-box or C/A-box motifs as high-affinity binding sites for STF1 and LONG HYPOCOTYL5 proteins. *Plant Physiol* **146**, 1862 (Apr, 2008).
273. Y. Meng, C. Shao, H. Wang, M. Chen, The regulatory activities of plant microRNAs: a more dynamic perspective *PLANT PHYSIOLOGY*, (2011).
274. C. G. Kawashima *et al.*, Interplay of SLIM1 and miR395 in the regulation of sulfate assimilation in Arabidopsis. *Plant J* **66**, 863 (Jun, 2011).

7 List of Figures and Tables

7.1 List of figures

7.2 List of figures supplementary

7.3 List of tables

7.4 List of tables supplementary

8 Abbreviations

Abbreviation	Full name
A. thaliana	*Arabidopsis thaliana*
-1D	dominant
1MO-ascorbigen	1-methoxyindol-3-ascorbigen
1MO-I3M	1 methoxy indol-3-ylmethyl glucosinolate
1MO-IAN	1-methoxyindol-3-acetonitrile
2,4 D	2,4-dichlorophenoxyacetic acid
4-MSOB-ITC	4-methylsulphinylbutyl isothiocyanate
4HO-I3M	4-hydroxy- indol-3-ylmethyl glucosinolate
4MO-ascorbigen	4-methoxyindol-3-ascorbigen
4MO-I3M	4 methoxy indol-3-ylmethyl glucosinolate
4MO-IAN	4-methoxyindol-3-acetonitrile
5-MSOP-ITC	5-methylsulphinylpentyl isothiocyanate
7-MSOH-ITC	7-methylsulphinylheptyl isothiocyanate
7MSOH	7-methylthioheptyl glucosinolate (glucosibarin)
8-MSOO-ITC	8-methylsulphinyloctyl isothiocyanate
8MSOO	8-methylsulfinyloctylglucosinolate (glucohirsutin)
A. tumefaciens	*Agrobacterium tumefaciens*
AA	amino acid
ABA	abscisic acid
ASML2	Activator of Spo^{min}::LUC2
AT-medium	Arabidopsis thaliana medium
ATP	adenosine triphosphate
BCA	bicinchoninic acid
BFA	brefeldin A
BSA	bovine serum albumin
CaMV	*Cauliflower mosaic virus*
CCT	CONSTANS, CONSTANS-like, TOC1
cDNA	complementary deoxyribonucleic acid
CDP-ME	4-diphosphocytidyl-2-C-methyl-D-erythritol
CE	capillary electrophoresis
CHI	chalcone isomerase
CHS	chalcone synthase
CoA	Coenzyme A
Col-*0*	Columbia
cp	chemoprotective
Ct	cycle threshold
CYPs	Cytochrome-P450-dependent mono-oxygenases
DAVID	Database for Annotation, Visualization and Integrated Discovery
DEPC	diethylpyrocarbonate
DFR	dihydroflavonol 4-reductase
DMAPP	dimethylallyl diphosphate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotide triphosphates
DPBS	Dulbecco's Phosphate Buffered Saline
DTT	dithiothreitol
DXP	1-deoxy-D-xylulose 5-phosphate
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EpRE	electrophile response element
ER	endoplasmatic reticulum
ERF	ethylene-responsive binding factor
ESI	electro spray ionization
et al.	et alia
EtBr	ethidium bromide

EtOH	ethanol
F3'H	cytochrome P450 monooxygenase
F3H	flavanone-3-hydroxylase
FAD	flavin adenine dinucleotide
FBS	fetal bovine serum
Fig	figure
FLS	flavonol synthase
FPP	farnesyl diphosphate
FTICR	Fourier transform ion cyclotron resonance
GAP	glyceraldehyde-3-phosphate
GC	gas chromatography
GFP	green fluorescent protein
GH3.5	Gretchen Hagen 3.5
GLL	GDSL lipase-like protein
GST	glutathione-S-transferase
GUS	Beta-glucuronidases
h	hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIG-1D	HIGH INDOLIC GLUCOSINOLATE1-1DOMINANT
HMG-CoA	3-hydroxy-3-methylglutaryl-CoA
HO-1	heme oxygenase-1
HPLC	high performance liquid chromatography
HY5	ELONGATED HYPOCOTYL 5
Hz	Hertz
I3C	indole-3-carbinol
I3M	indol-3-ylmethyl glucosinolate (glucobrassicin)
IAA	indole-3-acetic acid
IAN	indole-3-acetonitrile
IPP	isopentenyl diphosphate
ITC	isothiocyanate
JAL	jacalin-related lectin
kb	kilobases
KEAP1	Kelch-like ECH-associated protein 1
kV	kilo Volt
LB-medium	Lysogeny broth medium
LC	liquid chromatography
LIT	linear ion trap
Lux	luciferase
M	molar
m/z	mass-to-charge ratio
mA	milliampere
MALDI	matrix assisted laser desorption/ionization
MeOH	methanol
MEP	methylerythritol phosphate pathway
MES	2-(*N*-morpholino)ethanesulfonic acid
min	minute
MIR858a	MicroRNA858a gene
miR858a	mature microRNA858a
MMLV	Moloney Murine Leukaemia Virus
mRNA	messenger RNA
MS	mass spectrometry
MS-medium	Murashige & Skoog-medium
MTT	3-(4,5-dimethylthiazo2-yl)-2,5-diphenyltetrazolium bromide
N. benthamiana	*Nicotiana benthamiana*
n.d.	not determined
NAD	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NASC	Nottingham Arabidopsis Stock Center
ng	nanogram

nm	nanometer
Nrf2	nuclear factor-erythroid-2-related factor 2
OD	optical density
OPC8	3-oxo-2(2'-[*Z*]-pentenyl)cyclo-pentane-1-octanoic acid
OxIAA-Glc	oxindole-3-acetyl-glucose ester/glucoside
PAD3	phytoalexin deficient
PAL1	phenylalanine ammonia-lyase 1
PAL2	phenylalanine ammonia-lyase 2
PBP1	PYK10-binding protein 1
PCI	phenol/chloroform/isoamylalcohol
PCR	polymerase chain reaction
PEN2	PENETRATION 2
pGWB	gateway binary vector
QIT	quadruple ion trap
QR	NAD(P)H quinone oxidoreductase 1
QTOF	quadrupole-time of flight
RNA	ribonucleic acid
ROS	reactive oxygen species
rpm	revolutions per minute
RQ	relative quantification
RT	reverse transcription
s	second
SA	salicylic acid
SAM	shoot apical meristem
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulfate
SOB-medium	Super optimal broth medium
SOC-medium	Super optimal broth with Catabolite Repression
SU	sulfonylurea
SV-0	supervirulent
T-DNA	transfer- DNA
TAE	Tris-acetate-EDTA
TAIR	The Arabidopsis Information Resource
TAMARA	transposable element mediated activation tagging mutagenesis in
TE	Tris-EDTA
Tm	melting temperature
TOF	time of flight
tRNA	transfer ribonucleic acid
u	unit
UPLC	ultra performance liquid chromatography
UV	ultra violet
V	Volt
v/v	volume per volume
w/v	weight per volume
WOX2	WUSCHEL-related homeobox 2
YEB	Yeast extract and beef medium
µF	microfarad
µg	microgram
µL	microliter
µM	micromolar

9 Abstracts

9.1 Abstract

Many plant secondary metabolites have anti-cancerogenic and anti–mutagenic effects in higher organisms including humans. These compounds can induce the expression of phase II and detoxification enzymes, which inactivate carcinogens in mammalian cells. The genes encoding for phase II enzymes are regulated by electrophile response elements (EpRE) in their promotor region. In order to identify chemoprotective compounds, EpRE-based reporter systems have been developed in human and murine hepatoma cells, such as the HepG2-GFP and the Hepa1c1c7-Lux cells, respectively. The aim of this work was to identify novel chemoprotective plant secondary metabolites and the genes involved in their biosynthesis in *Arabidopsis thaliana* (*A. thaliana*). To this end, the effect of extracts of different activation tagged mutants (TAMARA lines) on the reporter cells was examined. In a previous screening with HepG2-GFP cells, 470 mutants had been selected out of 5,000 tested TAMARA lines. Here, these mutants were re-screened using the more sensitive Hepa1c1c7-Lux cell line. The screening and subsequent mapping of the T-DNA insertion sites led to the identification of 87 independent mutant lines. The relative expression of the genes flanking the insertion site was determined in comparison to the gene expression in wild-type plants. 32% of the detected genes encode for enzymes, 13% for transcription factors, 11% for transporters, 26% for proteins of unknown function, and 18% are genes belonging to other categories. The selection of mutants for further studies was based on the putative or known gene function and metabolite profiles from mass spectrometry analyses. Six different gene products that are overexpressed in independent mutant lines were selected for detailed characterization. These were encoded by the following genes: At3g09260 (ß-glucosidase PYK10), At3g54830 (a putative amino acid transporter), At1g63820 (an unknown protein), At5g55880/At5g55890 (an unknown protein), At4g27260 (AtGH3.5) and At1g71002 (MiR858a micro RNA).

9.2 Zusammenfassung

Viele pflanzliche Sekundärmetabolite besitzen antikanzerogene und antimutagene Eigenschaften in tierischen Organismen. Diese Wirkstoffe können in Säugerzellen die Expression von Phase II und detoxifizierenden Enzymen, welche zum Abbau von Zellgiften dienen, induzieren. „Electrophile response elements" (EpRE), lokalisiert in den Promotorregionen von Phase II-Enzymen, regulieren die Expression dieser Gene. Zur Identifizierung neuer chemoprotektiver Wirkstoffe wurden auf EpRE-basierende Reportersysteme in humanen (HepG2-GFP) und murinen Zelllinien (Hepa1c1c7-Lux) entwickelt. Zielsetzung dieser Arbeit war die Identifizierung neuer chemoprotektiver pflanzlicher Sekundärmetabolite in *Arabidopsis thaliana*. Außerdem sollten die an der Biosynthese dieser Pflanzenstoffe beteiligten Gene identifiziert und charakterisiert werden. Zu diesem Zweck wurde die Wirkung von methanolischen Extrakten aus verschiedenen Arabidopsis-Mutanten (TAMARA Linien) in den Reporterzellen untersucht. In früheren Experimenten mit HepG2-GFP-Zellen wurden bereits 470 aus insgesamt 5000 untersuchten TAMARA-Linien mit putativ chemoprotektiven Eigenschaften ausgewählt. In dieser Arbeit wurden diese Mutanten in den empfindlicheren Hepa1c1c7-Lux-Zellen erneut untersucht und ausgewählte Mutanten kartiert. Dies führte zur Identifizierung von 87 unabhängigen Linien. Die relative Genexpression der in direkter Nachbarschaft zur Insertionsstelle liegenden Gene wurde im Vergleich zur Genexpression in Wildtyppflanzen bestimmt. 32% der entdeckten Gene mit veränderter Expression kodieren für Enzyme, 13% für Transkriptionsfaktoren, 11% für Transporter, 26% für Proteine noch unbekannter Funktion und 18% gehören anderen Kategorien an. Sechs verschiedene Gene wurden für weitere Charakterisierungen ihrer Funktion auf Grundlage ihrer vermuteten Rolle in der Biosynthese von pflanzlichen Metaboliten bzw. aufgrund eines durch Massenspektrometrie ermittelten veränderten Metabolitenprofils ausgewählt. Bei diesen Genen handelt es sich um At3g09260 (ß-Glukosidase PYK10), At3g54830 (ein putativer Aminosäuretransporter), At1g63820 (ein unbekanntes Protein), At5g55880/ At5g55890 (ein unbekanntes Protein), At4g27260 (AtGH3.5) und At1g71002 (MiR858a mikroRNA).

10 Conferences

V. Ungewickell, T. Gigolashvili, S. Schmidt, H. Schmidt S. Clemens, D. Scheel and U.-I. Flügge (2011). Analysis of candidate genes identified in *Arabidopsis thaliana* activation-tagged lines with chemoprotective activities. PhD-Day, MPIZ, Köln (presentation).

Veronika Ungewickell, Tamara Gigolashvili, Stephan Schmidt, Holger Schmidt, Stephan Clemens, Dierk Scheel, Ulf-Ingo Flügge (2011). Analysis of candidate genes identified in *Arabidopsis thaliana* activation-tagged lines with chemoprotective activities. Botanikertagung der Deutschen Botanischen Gesellschaft, Berlin (Poster presentation).

V. Ungewickell, T. Gigolashvili, S. Schmidt, H. Schmidt, S. Clemens, D. Scheel and U.-I. Flügge (2010). Exploration of chemoprotective activities in *Arabidopsis thaliana* activation tagged lines. PhD-Day, MPIZ, Köln (Poster presentation).

V. Ungewickell, T. Gigolashvili, S. Schmidt, H. Schmidt, B. Beyene, S. Clemens, D. Scheel and U.-I. Flügge (2010). Exploration of chemoprotective activities in *Arabidopsis thaliana* activation tagged lines. 2nd Joint Retreat of PhD students in plant sciences, Köln (Poster presentation).

V. Ungewickell, T. Gigolashvili, S. Schmidt, H. Schmidt, B. Beyene, S. Clemens, D. Scheel and U.-I. Flügge (2010). Exploration of chemoprotective activities in *Arabidopsis thaliana* activation tagged lines. (BASIS – PROTECT) 10. GABI Status Seminar, Potsdam (Poster presentation).

Veronika Ungewickell, Tamara Gigolashvili, Stephan Schmidt, Holger Schmidt, Ulf-Ingo Flügge (2009). Exploration of chemoprotective activities in *Arabidopsis thaliana* activation tagged lines. Botanikertagung der Deutschen Botanischen Gesellschaft, Leibzig (Poster presentation).

T. Gigolashvili, S. Schmidt, H. Schmidt, V. Ungewickell, M. Humphry, S. Clemens, D. Scheel and U.-I. Flügge (2009). Exploration of novel chemoprotective compounds, which are beneficial to plant and human. (BASIS – PROTECT) 9. GABI Status Seminar, Potsdam (Presentation).

11 Acknowledgements & Formalities

11.1 Danksagung

Herrn Prof. Dr. U.-I. Flügge danke ich für die Bereitstellung eines interessanten und vielseitigen Themas und für die Möglichkeit meine Arbeit in seiner Gruppe durchzuführen.

Frau Prof. Dr. S. Waffenschmidt danke ich für die bereitwillige Übernahme des Korreferats.

Frau Prof. Dr. K. Schnetz danke ich für die freundliche Übernahme des Prüfungsvorsitzes.

Ganz besonders bedanke ich mich bei Dr. Tamara Gigolashvili für die Betreuung meiner Doktorarbeit und für die Ratschläge, Diskussionen und stetige Hilfsbereitschaft.

Bei Dr. S. Schmidt, Dr. H. Schmidt, Dr. C. Böttcher, B. Beyene, Prof. Dr. S. Clemens und Prof Dr. D. Scheel bedanke ich mich für die Zusammenarbeit und die Diskussionen auch im Rahmen der jährlichen Projekt-Treffen.

Der IMPRS Köln danke ich für die finanzielle Unterstützung.

PD Dr. Burkhard Becker danke ich für die Diskussion im Rahmen der Thesis Comittee.

Dr. Olof Persson und den Kollegen der IMPRS danke ich für die Diskussionen und die angenehme Atmosphäre.

Bei Dr. Henning Frerigmann, Natallia Ashykhmina, Sabine Wulfert, Sophia Titvinidze, Claudia Nothelle, Christina Uhlmann, Melanie Humphry, Lillian Marbaise, Christophe Eigen, Silvia Ebert, Eduard Hofsetz und Dunja Baatout bedanke ich mich für die stetige Hilfsbereitschaft, Unterstützung und Diskussionen.

Ein großes Dank geht an die Mitarbeiter der Gärtnerei, der Werkstatt und Siggi Werth.

Bei der gesamten AG-Flügge und den ehemaligen Mitgliedern der Arbeitsgruppe möchte ich mich für die angenehme Arbeitsatmosphäre, die Hilfsbereitschaft und der fachlichen Ratschläge sowie für gemeinsame Abende, Betriebsausflüge und weitere Aktivitäten bedanken.

Ganz besonders möchte ich mich bei meiner Familie für die große Unterstützung im Verlauf meiner Doktorarbeit bedanken.

11.2 Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen - , die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegt hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung der Universität zu Köln sind mir bekannt. Die von mir vorgelegte Arbeit ist von Prof. Dr. Ulf-Ingo Flügge betreut worden.

Veronika Ungewickell
Köln, April 2012

Lebenslauf

Veronika Ungewickell

Geburtsdatum: 06.07.1983
Geburtsort: Göttingen
Staatsangehörigkeit: Deutsch
Familienstand: Ledig

Schulausbildung:

1990-1993	Volksschule Planegg
1993-1995	Meramec Elementary School, Clayton, MO, USA
1995-1998	Wydown Middle School, Clayton, MO, USA
1998-2003	Sophienschule Hannover Abschluss: Abitur

Hochschulausbildung:

2003- 2006	**Leibniz Universität Hannover (Abschluss: Bachelor of Science)** Studiengang: Pflanzenbiotechnologie Bachelor-Arbeit: Überprüfung des Wirtspflanzenspektrums von *Potato Virus M* nach Agroinfektion und Nachweis von *Agrobacterium tumefaciens* Stamm C58C1 in *Nicotiana glutinosa* Pflanzenvirologie, Institut für Pflanzenkrankheiten und Pflanzenschutz
2006-2008	**Leibniz Universität Hannover (Abschluss: Master of Science)** Studiengang: Pflanzenbiotechnologie Master-Arbeit: Konstruktion eines kompletten, infektiösen cDNA Klones eines *Plum pox virus* Isolats aus der Sauerkirsche (PPV-SoC) Pflanzenvirologie, Institut für Pflanzenkrankheiten und Pflanzenschutz
Seit 2009	**Promotion** Universität zu Köln, Botanisches Institut, Prof. Dr. U.-I. Flügge